

WILEY

统计学精品译丛

金融数据分析导论

基于R语言

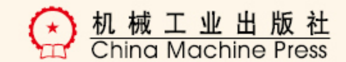
An Introduction to Analysis of Financial Data with R



（美）Ruey S. Tsay 著

芝加哥大学

李洪成 尚秀芬 郝瑞丽 译



本书由统计学领域著名专家所著，从基本的金融数据出发，讨论了这些数据的汇总统计和相关的可视化方法，之后分别介绍了商业、金融和经济领域中的基本时间序列分析和计量经济模型。作者通过实际操作的方法介绍金融数据分析，选择使用免费的R软件和实际案例来展示书中所讨论方法的实现。书中抽象理论和实际应用并重，读者既能从中轻松学习金融计量模型，也能了解它们在现实世界中的丰富应用。

贯通全书，各章节通过R图形以可视化的形式把讨论主题展现给读者，并以两个详细案例展示了金融中统计学的应用。本书有配套的网站（http://faculty.chicagobooth.edu/ruey.tsay/teaching/introTS），其中包含了书中涉及的R代码和额外的数据集供读者下载，通过这些读者可以创建自己的模拟分析，并检验对本书介绍的方法的理解程度。

本书是高年级本科生或研究生学习时间序列分析和商务统计学的优秀入门教材。对于希望进一步加强对金融数据和当今金融市场理解的研究人员以及商业、金融和经济领域的从业者，该书也是极佳的参考书。

作者简介



Ruey S. Tsay（蔡瑞胸） 美国芝加哥大学布斯商学院计量经济学与统计学的H.G.B. Alexander讲席教授，美国统计协会、数理统计学会以及英国皇家统计学会的会士，中国台湾“中央研究院”院士。他是《Journal of Forecasting》的联合主编，也是《Asia-Pacific Financial Markets》、《Studies in Nonlinear Dynamics and Econometrics》和《Metron》等期刊的副主编。Tsay教授在商务和经济预测、数据分析、风险管理以及过程控制等领域发表学术论文100多篇，还拥有美国专利“System and method for building a time series model (2005)”。

WILEY
www.wiley.com

客服热线：(010) 88378991 88361066
购书热线：(010) 68326294 88379649 68995259
投稿热线：(010) 88379604

数字阅读：www.hzmedia.com.cn
华章网站：www.hzbook.com
网上购书：www.china-pub.com

封面设计·杨宇梅

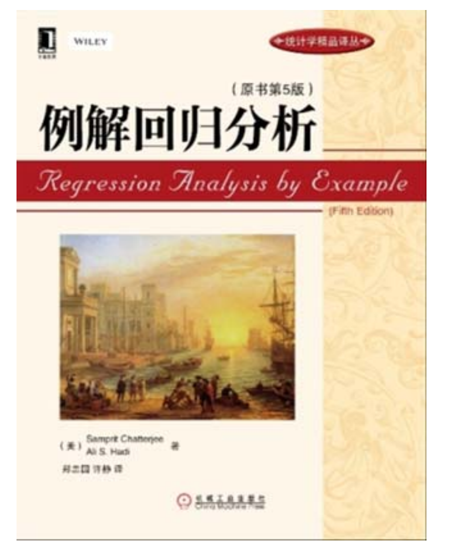
上架指导：经管/统计

ISBN 978-7-111-43506-8

9 787111 435068 >

定价：69.00元

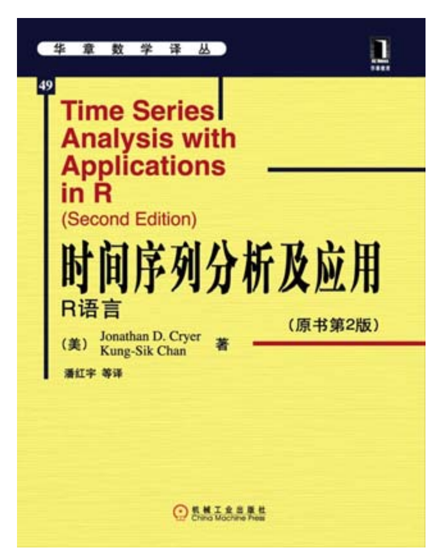
相关图书推荐



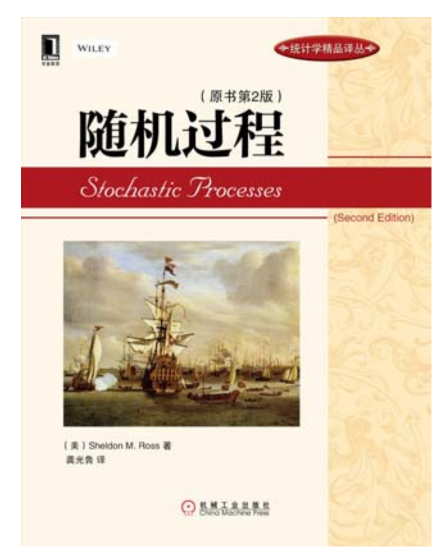
ISBN：978-7-111-43156-5



ISBN：978-7-111-40700-3



ISBN：978-7-111-32572-7



ISBN：978-7-111-43029-2

统计学精品译丛

金融数据分析导论

基于R语言

An Introduction to Analysis of Financial Data with R

（美） Ruey S. Tsay 著
芝加哥大学

李洪成 尚秀芬 郝瑞丽 译



图书在版编目（CIP）数据

金融数据分析导论：基于 R 语言 /（美）蔡瑞胸著；李洪成，尚秀芬，郝瑞丽译．—北京：机械工业出版社，2013.10

（统计学精品译丛）

书名原文：An Introduction to Analysis of Financial Data with R

ISBN 978-7-111-43506-8

Ⅰ．金…　Ⅱ．①蔡…　②李…　③尚…　④郝…　Ⅲ．金融－数据－分析－教材　Ⅳ．F830.41

中国版本图书馆 CIP 数据核字（2013）第 176528 号

本书向读者展示了可视化金融数据的基本概念，共有 7 章内容，涉及 R 软件、线性时间序列分析、资产波动率的不同计算方法、波动率模型在金融中的实际应用、高频金融数据的处理、用于风险管理的量化方法等．贯通全书，作者都是通过 R 图形以可视化的形式把讨论主题展现给读者，并以两个详细案例展示了金融中统计学的应用．

本书是高年级本科生或研究生阶段学习时间序列和商务统计学的优秀教材．对于希望进一步加强对金融数据和当今金融市场理解的研究人员以及商业、金融和经济领域的从业者，该书也是极佳的选择．

机械工业出版社（北京市西城区百万庄大街 22 号　　邮政编码　100037）

责任编辑：盛思源

印刷

2013 年 10 月第 1 版第 1 次印刷

186mm×240mm·19.75 印张

标准书号：ISBN 978-7-111-43506-8

定　　价：69.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：（010）88378991　88361066　　投稿热线：（010）88379604

购书热线：（010）68326294　88379649　68995259　　读者信箱：hzjsj@hzbook.com

推 荐 序

伟大的统计学家 George Box 有一句名言："所有的模型都是错误的，但其中有一些是有用的."这句话给出了现实中统计模型的现状. 现实中不确定性无处不在，确定性"正确"模型是不存在的. 如果 Box 在现在这个时代，他可能会增加一句"并且其中一些模型是危险的"，尤其对于金融领域的一些模型. 2008 年的金融危机，在某种程度上是由错误的金融模型造成的，既有模型过于简单的原因又有模型过于复杂的原因. 房地产经纪人和买家依赖于一个隐式模型，它表明价格已经在高位，并且还会继续上涨. 贷款人使用统计模型来对打包的按揭产品进行分析设计，这似乎可以奇迹般地降低风险. 然而最后的结果是灾难性的，在 5 年之后仍然可以感受到房地产泡沫的影响.

因此，如何进行有用的并且没有危险的金融分析呢？首先应该对金融数据有一个基本的理解，尤其是时间序列数据. 因为不确定性是主要的影响因素，例如，可以应用概率模型来描述资产收益率的频率分布. 本书给出了大量的时间序列模型，它们可以对数据进行描述、平滑和季节调整.

要成为一个统计分析专家，没有对实际数据的分析经验是绝不可能的. 蔡瑞胸教授在本书中给出了进行实际数据分析所需的数据和统计工具. 这里的统计分析工具是 R 软件，它是一款开源的统计软件包，可以和现在的商业软件包媲美. R 软件功能强大，免费并且有数千个用于完成特定任务的添加包. 蔡教授在书中给出了金融数据分析的几个关键添加包. 通过本书提供的数据集，你就可以应用 R 软件来学习金融时间序列的实际应用.

2013 年 6 月 17 日

美国统计教育学院，Statistics.com 在线课程网站总裁 Peter Bruce

Recommendation Preface

The great statistician George Box famously said "All models are wrong, but some are useful." This statement embodies the statistical approach to reality, in which uncertainty is everpresent, and a deterministic "truth" does not exist. Box might now add "and some models are dangerous" particularly those in the financial arena. The financial collapse of 2008 was caused, in part, by faulty financial models, both at a simple and a complex level. Real estate brokers and buyers relied upon an implicit model that said that prices have been going up, so they will continue to go up. Lenders used statistical models to engineer bundled mortgage products that magically seemed to reduce risk. The result was a disastrous real estate bubble whose effects were still being felt half a decade later.

So how do you conduct a financial analysis that is useful, and not dangerous? The first thing to focus on is a basic understanding of financial data-particularly time series data. Since uncertainty plays a large role, probabilistic models are used, for example, to portray the frequency distribution of asset returns. There are a number of time series models that describe, smooth, and seasonally adjust the data.

Expertise in statistical analysis cannot be obtained without getting the hands dirty in actual data and analysis. Ruey Tsay provides both the data and the statistical tool you will need to get your hands dirty. The tool is R-the popular open-source statistical software that is now on a par with commercial programs. It is powerful, free and has spawned the creation of thousands of R "packages" to accomplish specific tasks. Tsay provides pointers to some of the key packages for financial analysis. Together with the datasets provided in the book, you are equipped to learn the practical application of financial time series using R.

Statistics. com
The Institute for Statistics Education

译　者　序

本书是原作者继《Analysis of Financial Time Series》之后的又一本力作. 由于金融计量学和金融统计变得越来越重要，所以读者很迫切地需要这方面的资料. 而国内相关的教材目前还比较少，机械工业出版社及时引进了《金融数据分析导论》的中文版权，使得国内读者能够几乎和国外同步地学习该书的内容.

本书应用开源的 R 软件，结合具体的金融数据来讲解金融数据的分析方法和模型. 第 1 章首先向读者展示了可视化金融数据的基本概念、金融计量模型和它们在现实世界中的丰富应用. 第 2 章从线性时间序列模型开始，介绍常见的金融数据分析模型. 第 3 章介绍了 3 个金融数据分析的实际案例，读者可以从中学习线性时间序列模型在实际中的应用. 第 4 章介绍了波动率模型. 第 5 章介绍波动率模型在金融中的实际应用. 第 6 章介绍如何处理高频金融数据. 第 7 章介绍用量化方法进行风险管理，包括风险值和条件风险值.

对于实际的金融数据案例，本书应用免费的 R 软件给出了具体的分析过程和 R 代码. 读者可以按照书中的讲授，一步一步地进行实践，加深对本书内容的理解.

在本书的翻译过程中，我们得到了蔡瑞胸教授的帮助，他帮助译者澄清了一些问题. 中文版对原书中发现的错误进行了相应的更正. 本书第 1 章、第 3 章和第 5 章由尚秀芬翻译，第 2 章、第 4 章、第 6 章、第 7 章由李洪成完成，其中郝瑞丽协助翻译了第 2 章的部分内容. 全书由李洪成进行统一定稿. 本书的责任编辑盛思源为本书的出版付出了大量的劳动，由于她的认真校对和修订，本书才能及时得以出版. 另外，特别感谢明永玲编辑对本书翻译工作的大力支持和帮助. 本书的翻译还得到了上海金融学院的领导和同事的帮助和支持，在此一并表示衷心的感谢. 由于时间仓促，加之本人的精力和水平有限，翻译不当之处在所难免，请读者和同行指正.

李洪成

2013 年 6 月

前　言

我经常会被读者问到有关金融时间序列分析的问题：你能否让金融数据分析更简单些？我也经常被我的学生问道：如何简化实证研究工作，统计学和金融学究竟存在怎样的关系？这些都是很重要的问题，它们促使我编写这本入门教材.

为了简化实证数据分析，我决定使用 R 软件进行所有的分析. 这样做是基于下面几个原因. 第一，R 软件在大多数平台上都能免费使用. 第二，已有许多研究人员针对金融数据开发出很好的软件包，尤其是 RMetrics 提供了许多有用的软件包. 第三，R 软件添加包的功能正在飞速增加，并且这种趋势还会继续下去. 第四，我写过一些简单的 R 代码，用以在某些场合展示某些具体分析方法. 这些代码起两个作用. 其一，它们能满足我展示概念和方法的特殊需要. 其二，更重要的是，读者只要稍有一些 R 软件的使用经验就可以简单地修改这些代码和命令以适应他们的需要，从而简化他们的金融数据分析任务.

为了使计量经济学和统计学的理论更简单易懂，我尝试以一种简洁的方法来讲解，并应用大量实际案例来阐释这些内容. 本书共有 7 章，其中两章为案例分析. 这两章的案例揭示了统计学与金融学之间的联系. 其余章节用来帮助读者理解概念，获取金融数据分析的经验. 第 1 章介绍金融数据，并讨论了汇总统计量和数据可视化的知识. 此外，还介绍了 R 软件，这样读者就可以应用它来探索金融数据. 第 2 章介绍线性时间序列分析的基本内容，涵盖了商业、金融和经济学中常用的一些简单计量经济模型. 在保持行文简洁的同时，我尝试使该章内容尽量全面. 该章还介绍了指数平滑预测法和模型比较方法. 第 3 章介绍了 3 个案例学习（case study）. 其中使用的模型并不简单，但是它们可以帮助读者理解线性时间序列模型在实际应用中的价值和局限性. 第 4 章介绍资产波动率的不同计算方法，以及多种波动率模型. 其中讨论的方法包括应用日开盘价、收盘价、最高价和最低价的方法. 同样，我尝试使该章的内容尽量全面，同时避免过多的理论细节. 第 5 章介绍波动率模型在金融中的实际应用. 该章旨在帮助读者更好地理解波动率的期限结构，以及波动率在金融实践中的应用. 第 6 章介绍如何处理高频金融数据，包括价格变动、交易强度以及实际波动率的简单模型. 最后，第 7 章介绍用量化方法进行风险管理，包括风险值和条件风险值. 该章涵盖风险评估的重要计量经济学方法和统计学方法，包含基于极值理论和分位数回归的一些方法.

本书包含许多图表和示例，它们旨在简化金融数据分析的过程，使结果容易理解. 限于篇幅，不可避免地省略了一些重要内容. 本书涵盖的内容与《Analysis of Financial Time Series》一书有部分重叠，但是本书所用到的案例都是全新的.

借此机会，我感谢我的妻子. 如果没有她的关心和帮助，本书就无法完成. 同时我

也想感谢我的孩子，他们给予我很多灵感，帮助我编辑部分章节. 另外还有许多读者和学生不断给予我反馈和建议，他们的意见是弥足珍贵的. 最后，我想感谢 Steve Quigley、Jacqueline Palmieri 和他们的 Wiley 团队给予我的支持和鼓励.

本书网站：http://faculty.chicagobooth.edu/ruey.tsay/teaching/introTS.

R. S. T

伊利诺伊州芝加哥市

2012 年 10 月



目　　录

第1章　金融数据及其特征

近年来，数量分析方法在商业和金融市场上的重要性持续增加，因为我们有丰富的数据环境，经济和金融市场的数据比以前更加综合和完整．在许多国家，成百上千个变量的数据可以更系统、更精确地搜集，计算处理上的便利和统计软件包的使用使得对复杂的高维金融数据的分析成为可能．通过互联网，人们可以很容易地应用开源软件包下载公开的金融数据，比如R软件．所有这些软件的特性和功能免费公开，因而被广泛使用．

该书的目的是提供金融时间序列分析的基本知识，介绍分析金融数据的各种统计工具，获得各种计量经济学方法的金融数据应用经验．第1章，我们首先分析了本书有关的基本概念，通过实例引入了R软件．接下来，我们讨论R软件中用不同的方法可视化这些金融数据．第2章回顾了非线性时间序列的基本概念，如平稳性和自相关函数，介绍处理数据中序列依存关系的简单的线性模型．随后，讨论具有时间序列误差的回归模型、季节性调整以及单位根的非平稳性和长期记忆的过程．本章还讨论了用指数平滑法进行预测以及不同模型之间的比较．第3章对第2章介绍的模型进行具体应用，给出一些案例学习，有助于帮助读者更好地了解和分析金融数据、实证模型以及进行统计推断，同时也指出了线性时间序列模型在长期预测中的局限性．第4章主要讨论条件异方差模型的建模(即资产收益率的条件方差)．首先介绍刻画资产波动率随时间演变的计量经济模型．本章还讨论了不同的波动率模型的其他替代方法，包括使用资产的每日最高及最低价格来建模．第5章演示了不同波动率模型的应用，使用一些案例学习，给出波动率模型的建模步骤和过程，并对各种波动率模型的优缺点进行讨论，包括扩散限制的连续时间模型．第6章是关于高频金融数据的分析的．首先给出了高频数据的特性，分析高频数据相应的模型和方法，应用非同步交易和竞价反弹策略来研究股票收益率之间的序列相关问题．此外，还研究了交易间隔时间的动态性和分析交易数据的计量经济模型．特别地，我们讨论了应用逻辑线性回归模型和概率模型(probit model)来分析股票价格在连续交易中的运动过程．最后，使用日内对数收益率数据来研究实际波动率．第7章讨论金融头寸的风险度量及其在风险管理中的应用，引入了风险值和条件风险值来量化在金融头寸持有期间的各种风险，包括RiskMetrics模型、极值理论、分位数回归以及阈值峰值理论．

本书非常重视模型的应用和实证分析，每章都包含了很多实际的例子．在很多时候，金融时间序列的经验特征是计量经济模型发展的动因．对某些案例，本书网站上给出了某些具体分析的简单R脚本．每章后面的习题都给出了一些真实的数据可供使用．

1.1　资产收益率

大多数金融研究都是针对资产收益率，而不是资产价格．Campbell等(1997)给出了使用资产收益率的两个主要原因．首先，对于一个普通的投资者来说，资产收益率代表一个完全的、尺度自由的投资机会的总结和概括．其次，资产收益率序列比价格序列更容易处

理，前者有更好的统计特性．然而，资产收益率有多种不同的定义．

设 P_t 是 t 时刻的资产价格，下面给出本书中用到的一些资产收益率的定义．暂时假定该过程资产不支付红利．

2

单期简单收益率

假设投资者在一个周期内拥有某种资产，从第 $t-1$ 天到第 t 天，其**简单毛收益率**为：

$$1+R_t=\frac{P_t}{P_{t-1}} \text{ 或 } P_t=P_{t-1}(1+R_t) \tag{1-1}$$

相对应的**单期简单净收益率**(simple net return)或**简单收益率**(simple return)为：

$$R_t=\frac{P_t}{P_{t-1}}-1=\frac{P_t-P_{t-1}}{P_{t-1}} \tag{1-2}$$

表 1-1 给出了 2011 年 12 月苹果公司股票每日收盘价．从表 1-1 中可知，从 2011 年 12 月 8 日到 12 月 9 日，持有苹果公司股票每日的总收益率为 $1+R_t=393.62/390.66\approx 1.0076$．相应的每日简单收益率为 0.76%，即(393.62−390.66)/390.66=0.76%．

多期简单收益率

假设从第 $t-k$ 天到第 t 天，这 k 个周期内持有某种资产，则 ***k* 期简单毛收益率**为：

$$\begin{aligned}1+R_t[k]&=\frac{P_t}{P_{t-k}}=\frac{P_t}{P_{t-1}}\times\frac{P_{t-1}}{P_{t-2}}\times\cdots\times\frac{P_{t-k+1}}{P_{t-k}}\\&=(1+R_t)(1+R_{t-1})\cdots(1+R_{t-k+1})\\&=\prod_{j=0}^{k-1}(1+R_{t-j})\end{aligned}$$

这样，k 期简单毛收益率是其包含的这 k 个单期简单毛收益率的乘积，称为**复合收益率**(compound return)．k 期简单净收益率为 $R_t[k]=(P_t-P_{t-k})/P_{t-k}$．

为了说明这一点，表 1-1 给出了苹果公司股票的每日收盘价．由于 2011 年 12 月 2 日和 9 日是星期五，所以每周的股票简单毛收益率为 $1+R_t[5]=393.62/389.70\approx 1.0101$，即每周的简单收益率为 1.01%．

表 1-1　苹果公司股票从 2011 年 12 月 2 日到 12 月 9 日的每日收盘价

日期	12/02	12/05	12/06	12/07	12/08	12/09
价格($)	389.70	393.01	390.95	389.09	390.66	393.62

在实际中，确切的时间区间对讨论和比较收益率是非常重要的(例如月收益率还是年收益率)．若时间区间没有给出，这里隐含的假定时间区间为一年．如果持有资产的期限为 k 年，则(平均)年度化收益率定义为

$$\text{年度化的}\{R_t[k]\}=\left[\prod_{j=0}^{k-1}(1+R_{t-j})\right]^{1/k}-1$$

这是由它所包含的 k 个单期简单毛收益率几何平均得到的，可用下式计算：

$$\text{年度化的}\{R_t[k]\}=\exp\left[\frac{1}{k}\sum_{j=0}^{k-1}\ln(1+R_{t-j})\right]-1$$

其中 $\exp(x)$表示指数函数，$\ln(x)$是正数 x 的自然对数．因为算术平均值比几何平均值计算起来容易，并且单期收益率一般很小，所以我们可用一阶泰勒(Taylor)展开来近似表示年度化的收益率，则有

$$\text{年度化的}\{R_t[k]\} \approx \frac{1}{k}\sum_{j=0}^{k-1} R_{t-j} \tag{1-3}$$

然而，在有些应用中，式(1-3)的近似精确度可能不够．

连续复利收益率

在引进连续复利收益率之前，先讨论复利收益率的效果．假定银行存款的年利率为10%，最初的存款为1美元．如果银行每年支付一次利息，则一年后存款的净值变为 $1\times(1+0.1)=1.1$ 美元．如果该银行半年付息一次，则六个月的利息率为 $10\%/2=5\%$，一年后净值为 $1\times(1+0.1/2)^2=1.1025$ 美元．通常，如果银行一年付息 m 次，则每次支付的利率为 $10\%/m$，一年后存款的净值变成了 $1\times(1+0.1/m)^m$ 美元．表1-2给出了年利率为10%时，一些常用的时间间隔下存款1美元的净值结果．尤其是，净值 $\exp(0.1)\approx 1.1052$ 美元，这个值是**连续复利的结果**(result of continuous compounding)．我们可以清楚地看到复利的效果．

表 1-2　收益率复利效果的演示(期限为1年，年利率为10%)

类型	支付次数	每期利率	净值
一年	1	0.1	$1.100 00
半年	2	0.05	$1.102 50
季度	4	0.025	$1.103 81
每月	12	0.0083	$1.104 71
每周	52	0.1/52	$1.105 06
每天	365	0.1/365	$1.105 16
连续	∞		$1.105 17

通常，连续复利的资产净值 A 可以表示为

$$A = C\exp(r\times n) \tag{1-4}$$

其中 r 是年利率，C 是初始资本，n 是持有资产的年数．由式(1-4)，则有

$$C = A\exp(-r\times n) \tag{1-5}$$

这里，C 是从现在开始 n 年后价值为 A 的资产的现值．假定 r 是用连续复利表示的年利率．

连续复合收益率

资产的简单毛收益率的自然对数称为**连续复合收益率**或**对数收益率**(log-return)：

$$r_t = \ln(1+R_t) = \ln\frac{P_t}{P_{t-1}} = p_t - p_{t-1} \tag{1-6}$$

其中，$p_t=\ln(P_t)$，与简单净收益率 R_t 相比，连续复合收益率 r_t 有很多优点．首先，对多期收益率，我们有

$$\begin{aligned} r_t[k] &= \ln(1+R_t[k]) = \ln[(1+R_t)(1+R_{t-1})\cdots(1+R_{t-k+1})] \\ &= \ln(1+R_t)+\ln(1+R_{t-1})+\cdots+\ln(1+R_{t-k+1}) \\ &= r_t + r_{t-1} + \cdots + r_{t-k+1} \end{aligned}$$

这样，连续复合多期收益率是它所包含的连续复合单期收益率之和．其次，对数收益率具有更容易处理的统计特性．

为了说明这一点，我们再次考虑表1-1中苹果公司股票的日收盘价格．从12月8日到

12月9日的日对数收益率为 $r_t=\log(393.62)-\log(390.66)\approx0.75\%$，从12月2日到12月9日的周对数收益率为 $r_t[5]=\log(393.62)-\log(389.70)\approx1.00\%$. 显而易见，周对数收益率是该周的5个日对数收益率之和.

资产组合收益率

若一个资产组合由 N 项资产组成，则该资产组合的简单净收益率是它所包含的各项资产的简单净收益率的加权平均，其中每个资产的权重是该资产的价值占资产组合总价值的百分比. 设 p 是一个资产组合，它在第 i 项资产上的权重为 w_i，那么 p 在 t 时刻的简单收益率为 $R_{p,t}=\sum_{i=1}^{N}w_iR_{it}$，其中 R_{it} 是组合中第 i 个资产的简单收益率.

然而，资产组合的连续复合收益率没有上述方便的性质. 如果简单收益率 R_{it} 的绝对值很小，我们有 $r_{p,t}\approx\sum_{i=1}^{N}w_ir_{it}$，其中 $r_{p,t}$ 是该组合在 t 时刻的连续复合收益率. 经常用这种近似来研究资产组合的收益率.

3~5

红利支付

如果一项资产周期性地支付红利，则我们需要修改资产收益率的定义. 设 D_t 是一项资产在第 $t-1$ 天和第 t 天之间支付的红利，P_t 是该资产在第 t 个周期末的价格. 这样，红利并没有包含在 P_t 中. 则 t 时刻的简单净收益率和连续复合收益率分别为

$$R_t=\frac{P_t+D_t}{P_{t-1}}-1,\quad r_t=\ln(P_t+D_t)-\ln(P_{t-1})$$

超额收益率

一项资产在 t 时刻的超额收益率是该项资产的收益率与某项参照资产的收益率之差. 参照资产通常是无风险的，如美国短期国债收益率. 简单超额收益率和对数超额收益率分别定义为

$$Z_t=R_t-R_{0t},\quad z_t=r_t-r_{0t}\tag{1-7}$$

其中，R_{0t} 和 r_{0t} 分别是该参照资产的简单收益率和对数收益率. 在金融学文献中，超额收益率被认为是一个套利投资组合的盈利. 在这个投资组合中，对某资产持有多头头寸而对其参照资产持有空头头寸，且初始净投资额为零.

注记：多头金融头寸是指持有某资产. 空头头寸是指卖出不属于自己的资产，这需要通过从已购买该资产的投资者那里借入资产来实现. 随后的某天，卖空者有义务买进与借入时完全相同数量的份额来偿还借出者. 偿还时要求相等数量的份额，而不是相等数量的美元，因此卖空者会由于该资产价格的下跌而获利. 如果在空头持有期间该资产支付现金红利，这些红利将支付给做空合约的买方. 卖空者必须从自己的资源里支付相应的现金红利来补偿借出者. 换句话说，空头方有义务支付所借资产的现金红利给资产的借出方. ■

关系小结

简单收益率 R_t 与连续复合对数收益率 r_t 的关系是

$$r_t=\ln(1+R_t),\quad R_t=e^{r_t}-1$$

如果收益率 R_t 与 r_t 用百分数表示，则有

$$r_t = 100\ln\left(1+\frac{R_t}{100}\right),\quad R_t = 100(\mathrm{e}^{r_t/100}-1)$$

把收益率进行时间累加，则有

$$1+R_t[k] = (1+R_t)(1+R_{t-1})\cdots(1+R_{t-k+1})$$

$$r_t[k] = r_t + r_{t-1} + \cdots + r_{t-k+1}$$

如果连续复合年利率为 r，则资产现值与资产的未来价值之间的关系为

$$A = C\exp(r\times n),\quad C = A\exp(-r\times n)$$

例 1.1 若某项资产的月对数收益率为 4.46%，则相应的月简单收益率是 100[exp(4.46/100)−1]=4.56%. 同样，若某项资产在一个季度内的月对数收益率分别为 4.46%、−7.34% 和 10.77%，则该资产季度的对数收益率为 (4.46−7.34+10.77)%=7.89%. ■

图 1-1 显示了 IBM 股票的日简单收益率及其对数收益率的时序图，从 2001 年 1 月 2 日到 2010 年 12 月 31 日，共有 2515 个观测值. 从图 1-1 可知，和简单收益率相比，对数收益率更简单. 事实上，简单收益率和对数收益率之间的相关系数为 0.9997. 这是可以理解的，因为当 x 接近零时，$\log(1+x)\approx x$. 在样本持续期内，IBM 股票的日简单收益率的数额很小.

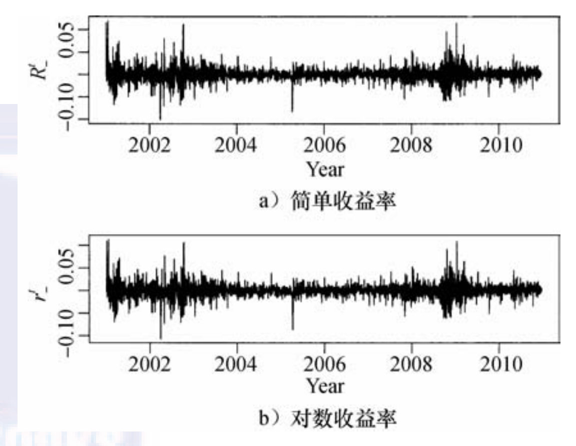


图 1-1 IBM 股票从 2001 年 1 月 2 日到 2010 年 12 月 31 日的日收益率时序图

1.2 债券收益和价格

6～7

债券是一种金融工具，在到期日，向债券的持有者支付票面价值(或面值). 有些债券根据**票面利率**(coupon payment)定期支付利息. 而零息债券则不需要定期支付利息. 债券收益率是持有债券至到期日，投资者将收取的回报. 在金融上，有几种类型的常用债券. 这些债券的收益率是指当期收益率和到期收益率(Yield To Maturity，YTM).

当期收益率

当期收益率是指每年支付给投资者回报的百分数. 从数学上，我们有

$$\text{当期收益率} = \frac{\text{支付的年度利息额}}{\text{债券的市场价格}}\times 100\%$$

例如，一个投资者购买债券支付 90 美元，债券面值(也称为**面值**(par value))为 100 元，债券的票面利率为每年 5%，则该债券当期收益率为

$$c_t = (0.05\times 100)/90\times 100\% = 5.56\%$$

这里，我们用下标 t 表示收益率通常随时间变化. 从定义可知，当期收益率不包含任何投资过程中的资本收益或损失. 对于零息债券，收益率的计算方法如下：

$$\text{当期收益率} = \left(\frac{\text{面值}}{\text{购买价格}}\right)^{1/k} - 1$$

这里，k 指用年度表示的到期时间. 举例来说，如果投资者购买零息债券，支付 90 美元，其面值为 100 美元，该债券在 2 年内到期，则该债券的收益率为 $c_t=(100/90)^{1/2}-1=5.41\%$.

到期收益率

当期收益率没有考虑货币的时间价值，因为它没有考虑投资者未来收到的债券利息的当前价值. 因此，一个较为常用的债券投资收益是到期收益率(YTM). 然而，到期收益率的计算更复杂. 简单地说，到期收益率是通过将债券的价格与将来所有的现金流入的现值相等而计算出的收益率. 在购买日和到期日之间，假设该债券的投资者收到 k 期利息支付. y 为债券的到期收益率，P 为债券的价格，则有

$$P=\frac{C_1}{1+y}+\frac{C_2}{(1+y)^2}+\cdots+\frac{C_k+F}{(1+y)^k}$$

这里，F 为债券的面值，C_i 是第 i 期的利息支付. 假定每年的债券票面利率为 α，每年支付次数为 m，到期时间为 n 年，则债券利息每期的现金流入为 $F\alpha/m$，支付期数为$k=mn$. 债券价格和到期收益率计算如下

$$\begin{aligned}P&=\frac{\alpha F}{m}\left[\frac{1}{1+y}+\frac{1}{(1+y)^2}+\cdots+\frac{1}{(1+y)^k}\right]+\frac{F}{(1+y)^k}\\&=\frac{\alpha F}{my}\left[1-\frac{1}{(1+y)^k}\right]+\frac{F}{(1+y)^k}\end{aligned}$$

8

下表给出了债券价格和到期收益率的一些结果，这里假设 $F=100$ 美元，票面利率为 5%，每半年支付一次，到期时间为 3 年.

到期收益率(%)	半年收益率(%)	债券价格(美元)
6	3.0	97.29
7	3.5	94.67
8	4.0	92.14
9	4.5	89.68
10	5.0	87.31

从表中可以看到，随着债券到期收益率的增加，债券价格在下降. 换句话说，到期收益率和债券价格成反比. 在实践中，通过观测到的债券价格，可以计算出到期收益率. 一般不容易得到精确的解，但我们可以获得一些精确的近似解. 例如，在前面的表中，购买债券的投资者支付的价格为 94 美元，则到期收益率在 7%～8%之间. 通过试验-更正错误的试验，得到投资者每年到期收益率约为 7.26%. 许多金融机构提供在线程序来计算债券到期收益率及其价格，例如，Fidelity 投资公司.

到期收益率(%)	半年收益率(%)	债券价格(美元)
7.1	3.55	94.41
7.2	3.6	94.16
7.3	3.65	93.90
7.25	3.625	94.03
7.26	3.63	94.00

美国政府债券

美国政府发行各种债券，为其债务进行融资．这些债券包括短期国债、中期国债和长期国债．下面给出这些债券的简单描述．

- 短期国债(Treasury Bills，T-Bills)在一年或一年以内到期．在到期前，它们不支付任何利息，按照面值来折价出售，从而产生正的到期收益率．常用的期限有28天(1个月)、91天(3个月)、182天(6个月)和364天(1年)．最低的购买价格是100美元．短期国债折现收益率(相当于用年表示的当期收益率)计算如下：

$$\text{折现收益率}(\%) = \frac{F-P}{F} \times \frac{360}{\text{到期持有天数}} \times 100(\%)$$

  这里，F和P分别表示债券的面值和购买价格．美国财政部每星期四宣布13周和26周国债的发行数量，接下来的周一举行拍卖，周四进行结算．每周一宣布4周国债发行数量，周二拍卖，周四结算．第4个星期四宣布52周国债的发行数量，接下来的周二拍卖，周四结算．

- 中期国债(Treasury Notes，T-Notes)在1～10年内到期．每6个月支付一次利息，面值为1000美元．这些票据在二级市场上以面值的32进制的百分数来报价．例如，报价为95：08，这说明该债券以折价进行交易，即价格为(95＋8/32)×1000/100＝952.5美元．当我们讨论美国政府债券市场的时候，10年期国债已经成为报价频率和安全性最高的债券．详见下一节芝加哥期权交易所(Chicago Board Options Exchange，CBOE)的10年期债券．图1-5和图1-7分别列出了美国10年期政府债券日收益率及其收益率的时序图．
- 长期国债(Treasury Bonds，T-Bonds)有更长的到期期限，从20～30年不等．每6个月支付一次利息，政府一般发行30年期的国债．从2001年10月31日开始，30年期债券暂停了4年零6个月．在2006年2月政府开始重新发行，每季度发行一期．

1.3 隐含波动率

股票期权是一种金融合约．股票A的看涨期权是这样一种合约，即合约持有者在给定的期间内以约定的价格购买一定数量股票A的权利，而不是义务．同时，股票A的看跌期权是这样的一种合约，即合约的持有者在给定的期间内以约定的价格卖出一定数量股票A的权利，而不是义务．这里约定的价格称为**执行价格**(strike price)，给定的时间期间称为**到期日**(time to maturity)．在美国，每份股票期权通常包含有100股股票．期权在期权市场进行交易，如芝加哥期权交易所(CBOE)．期权有很多种类型．常用的有欧式期权，即只有在到期日才可行使其权利．美式期权，可以在到期日及到期日之前的任何时刻行使其权利．详细的介绍参见Hull(2011)．如果期权在立即执行时给其持有人的现金流为正，我们称这种期权为**价内期权**(in-the-money)．如果期权在执行时，给持有人的现金流为负，我们称其为**价外期权**(out-of-the-money)．最后，如果期权在执行时给持有人的现金流为零，我们称它为**平价期权**(at-the-money)．

9

股票期权价格的影响因素有很多，比如执行价格、无风险利率、当前股价以及股票的波动率．例如，著名的 Black-Scholes 模型．该模型在股票价格服从几何布朗运动的前提下，推导出其解析解．本章中，只要注意到 Black-Scholes 模型中唯一的一个不能直接观测到的因子是股票的波动率就可以了．这里的波动率指股票价格的条件标准差．在实践中，我们可以用观测到的股票价格和 Black-Scholes 模型来反向推导出其波动率，这个波动率称为**隐含波动率**(implied volatility)．类似于债券到期收益率的实现过程，隐含波动率可以近似得到．

10

大家都熟知的隐含波动率是芝加哥期权交易所(CBOE)的波动率指数(Volatility Index，VIX)．该指数最初在 1993 年推出，目的是通过 S&P 100 平价指数期权价格来度量 30 天隐含波动率的市场期望值．该指数在 2003 年由芝加哥期权交易所修订，修订后的指数反映了期望波动率新的度量方法．现在是 S&P 500 指数(SPX)和估计的期望波动率的加权平均，权重是一个很大范围的执行价格下的 S&P 500 指数看涨期权及看跌期权的价格．详细信息请参阅 CBOE VIX 白皮书．新的 VIX 通常被视为市场恐慌情绪的影响因素，在金融市场上发挥着重要的作用．事实上，波动率指数的期货和期权也在芝加哥期权交易所进行交易．

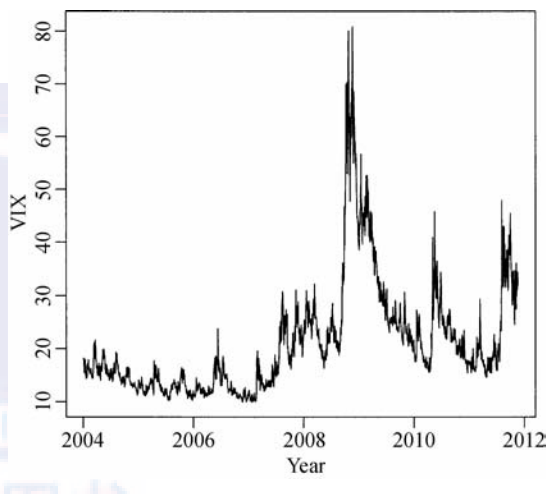


图 1-2　从 2004 年 1 月 2 日到 2011 年 11 月 21 日芝加哥期权交易所波动率指数的时序图

图 1-2 显示了从 2004 年 1 月 2 日到 2011 年 11 月 21 日修订后的 VIX 指数的时序图，共 1988 个观测值．从图 1-2 中可知，在 2008 年年底和 2009 年年初，金融市场的波动性非常高．在 2011 年市场的波动性也很高．在后面的章节，我们将分析波动率指数．关于资产波动率更详细的信息，请参阅本书的第 4 章．

1.4　R 软件包及其演示

在研究金融数据的真实例子之前，我们先简要地介绍本书中广泛应用的 R 软件．当数据分析中需要用到相关的 R 添加包及其命令时，将给出相应的介绍．我们的目的是使实证分析尽可能容易，这样读者可以自己运行程序来得到本书给出的结果．

R 是一款免费软件，可以从 http://www.r-project.org 下载．R 可以在很多操作系统上运行，包括 Linux、MacOS X 和 Windows．在 R 软件的网站上我们可以点击“CRAN”，选择附近的 CRAN 镜像，下载并安装该软件并选择其添加包．最简单的安装方法是按照在线指导，使用默认选项来完成．由于 R 是一个开放源代码软件，所以它包含了研究者开发的数百个添加包，以供世界各地的研究人员进行各种统计分析和应用．对于金融时间序列分析，Diethelm Wuertz 博士和他的同事开发了很多有用的软件包，包括 fBasics、fGarch 和 fPortfolio.

本书中我们使用这些添加包的很多功能. 我们还可以使用 R 其他的添加包，这些添加包功能更强大和易于使用，如 R 中用于极值分析的 evir 添加包. 关于 R 安装及命令的更多信息，可在本书的网页或作者的教学网页上找到. 现在有多本关于 R 的入门书籍，例如，可以参见 Adler(2010)和 Crawley(2007). R 命令区分大小写，必须完全遵守.

1.4.1 R 软件包的安装

使用默认选项安装 R 软件后，将在计算机桌面上创建一个 R 图标. 只需双击该 R 图标，即可启动 R 程序. 对于 Windows 系统，将会给出一个带有命令菜单和 R 控制台的 `RGui` 窗口. 要安装 R 添加包，可以单击菜单“程序包”，选择“安装程序包”来选择需要安装的添加包. 这时会弹出一个窗口，询问用户选择 R 镜像(和前面提到过的 R 安装相类似). 选择好镜像后，会弹出另一个窗口，里面包含所有可用的添加包，可以单击所需的添加包进行安装.

安装完成添加包后，可以通过单击“程序包”菜单的“加载程序包”子命令，就可以将该添加包加载到 R 中. 单击该子命令后，会弹出一个窗口，它列出所有已经安装的添加包，供用户进行选择. 另一种加载添加包的方法是使用 `library` 命令. 详见下面讨论中的演示.

1.4.2 Quantmod 软件包

首先，我们看看常用的 R 添加包，可以从一些开放源直接下载金融数据，这些网站包括雅虎财经、谷歌财经和圣路易斯联邦储备银行的联邦储备经济数据库(FRED). 强烈建议安装 Jeffry 开发的 quantmod 添加包. 该添加包还需要安装 TTR、xts 和 zoo 三个额外的添加包.

11~12

安装完毕以后，quantmod 添加包允许用户与互联网连接，用户可以使用一些命令符来访问雅虎和谷歌财经的日股票数据，使用这些序列名称来访问来自联邦储备经济数据库超过 1000 个经济和金融时间序列数据. 命令是 `getSymbols`. 该添加包有很好的功能，例如，命令 `chartSeries` 可以直接画出这些序列的收盘价和交易量的时序图. 这两个命令的默认选项对于分析金融时间序列来说是足够的. 每个人都可以使用其子命令进一步扩展该添加包的功能，如在命令 `getSymbols` 中设定收益率的时间跨度. 有兴趣的读者可以查阅这些命令的相关描述. 这里，我们提供一个简单示例. 图 1-3 显示了苹果公司股票从 2007 年 1 月 3 日至 2011 年 12 月 2 日的日收盘价及成交量的时序图. 该图显示了最近可

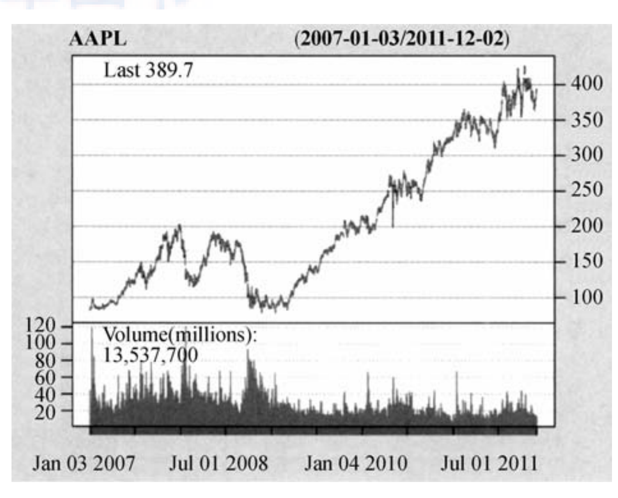


图 1-3 从 2007 年 1 月 3 日至 2011 年 12 月 2 日苹果公司股票的日收盘价及成交量的时序图

观测数据的价格和交易量. chartSeries 中的子命令 theme="white"可用来设置时序图的背景，默认为黑色. 图 1-4 显示了从 1948 年 1 月至 2011 年 11 月美国月失业率的时序图. 图 1-5 显示了从 2007 年 1 月 3 日至 2011 年 12 月 2 日 10 年期美国国债日利率的时序图. 这些利率来自芝加哥期权交易所(CBOE)，数据来自雅虎财经网站. 由于没有成交量，所以在 chartSeries 中，子命令 TA= NULL 常常用来表示时序图中忽略交易量. 命令 head 和 tail 分别给出数据集的前 6 行数据和最后 6 行数据.

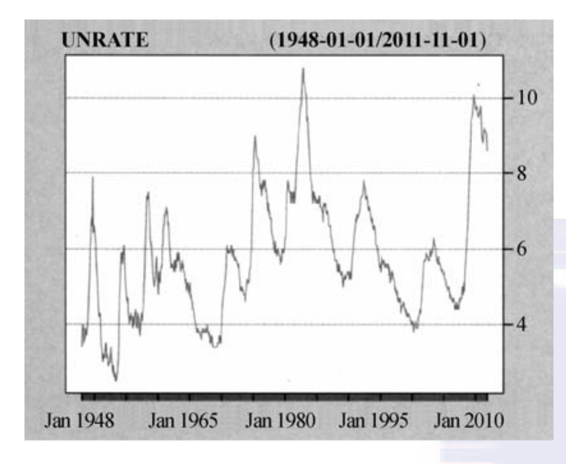


图 1-4 从 1948 年 1 月至 2011 年 11 月美国月失业率的时序图

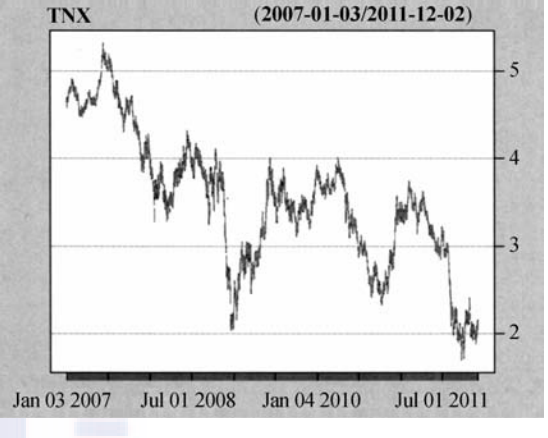


图 1-5 从 2007 年 1 月 3 日至 2011 年 12 月 2 日美国 10 年期国债日利率的时序图

带有 quantmod 添加包的 R 演示(本书中的 R 输出经过了排版编辑，“>”表示命令提示符、%表示其后为注释内容)

```
> library(quantmod)   % Load the package
> getSymbols("AAPL")  % Download daily prices of Apple stock from Yahoo
[1] "AAPL" % I ran R on 2011-12-03 so that the last day was 12-02.
> dim(AAPL)   % (dimension): See the size of the downloaded data.
[1] 1241    6
> head(AAPL)  % See the first 6 rows of the data
              Open      High     Low      Close      Volume   Adjusted
2007-01-03    86.29     86.58    81.90    83.80    44225700      83.80
2007-01-04    84.05     85.95    83.82    85.66    30259300      85.66
....
2007-01-10    94.75     97.80    93.45    97.00   105460000      97.00
> tail(AAPL)   % See the last 6 rows of the data
            Open      High     Low      Close      Volume   Adjusted
2011-11-25  368.42    371.15   363.32   363.57    9098600     363.57
.....
2011-12-01  382.54    389.00   380.75   387.93   13709400     387.93
2011-12-02  389.83    393.63   388.58   389.70   13537700     389.70

> chartSeries(AAPL,theme="white")   % Plot the daily price and volume
% The subcommand theme is used to obtain white background of the plot.
> chartSeries(AAPL)%Not shown giving the same plot with black background.
```

```
% The next command specifies the data span of interest
> getSymbols("AAPL",from="2005-01-02", to="2010-12-31")
[1] "AAPL"
> head(AAPL)
           AAPL.Open AAPL.High AAPL.Low AAPL.Close AAPL.Volume Adjusted
2005-01-03     64.78     65.11    62.60      63.29    24714000    31.65
2005-01-04     63.79     65.47    62.97      63.94    39171800    31.97
 ......
> getSymbols("UNRATE",src="FRED")%Download unemployment rates from FRED.
[1] "UNRATE"
> head(UNRATE)
           UNRATE
1948-01-01    3.4
1948-02-01    3.8
 ......
1948-06-01    3.6
> chartSeries(UNRATE,theme="white")  % Plot monthly unemployment rates
% The subcommand "src" is used to specify the data source.
% The default is Yahoo.
> getSymbols("INTC",src="google")  % Download data from Google.
[1] "INTC"
> head(INTC)
           INTC.Open INTC.High INTC.Low INTC.Close INTC.Volume
2007-01-03     20.45     20.88    20.14      20.35    68665100
2007-01-04     20.63     21.33    20.56      21.17    87795400
.....
2007-01-10     21.09     21.62    21.03      21.52    75522200

> getSymbols("^TNX") % Download CBOE 10-year Treasures Notes
[1] "TNX"
> head(TNX)
           TNX.Open TNX.High TNX.Low TNX.Close Volume Adjusted
2007-01-03     4.66     4.69    4.64      4.66     0     4.66
2007-01-04     4.66     4.66    4.60      4.62     0     4.62
.....
2007-01-10     4.67     4.70    4.66      4.68     0     4.68
> chartSeries(TNX,theme="white",TA=NULL) % Obtain plot without volume.
```

1.4.3 R的基本命令

启动R以后，要做的第一件事是设置工作目录．所谓工作目录，是指我们存放数据集和输出结果的计算机目录．可以用两种途径实现．第一种方法是单击“文件”菜单的“改变工作目录……”子命令，弹出一个窗口，此时可以选择所需要的目录．第二种方法是通过R控制台，使用命令 setwd 来设置工作目录．具体可以参考下面的演示．

R是一款面向对象的软件，可以处理多种类型的对象．本书中我们不需要研究R中对象的详细信息，用到时将给出解释．可以说，R允许给变量赋值，通过变量名称来引用这些值．赋值运算符是“＜－”，也可以使用“＝”．比如，“x＜－10”表示把10这个数值赋给变量“x”，这里R把“x”作为一个实数序列，其第一个元素为10．有几种方法可以将数据加载到R的工作空间，具体方式取决于数据的格式．对于简单的纯文本数据，其命令是 read.table．对于 .csv 文件，命令是 read.csv．可以用单引号或者双引号来指定数据文件，详见R代码演

示. R 将数据作为对象，通过为它们指定的名称来引用这些数据. 对于上述两个数据加载命令，R 将数据存储在矩阵框中. 还可以使用命令 dim(即，维度)来查看数据集的大小. 最后，R 中的基本操作和我们通常使用的软件类似，可以使用命令 q()选择退出.

R 代码演示

```
> setwd("C:/Users/rst/book/introTS/data") % Set my working directory
> library(fBasics) % Load package
> x <- 10  % Assign value, here "x" is a variable.
> x  % See the value of x.
[1] 10 % Here [1] signifies the first element.
> 1 + 2 % Basic operation: addition
[1] 3
> 10/2 % Basic operation: division
[1] 5
% Use * and ^ for multiplication and power, respectively.
% Use log for the natural logarithm.
> da=read.table('d-ibm-0110.txt',header=T) % Load text data with names.
> head(da) % See the first 6 rows
      date    return
1 20010102 -0.002206
2 20010103  0.115696
 ....
6 20010109 -0.010688
> dim(da) % Dimension of the data object "da".
[1] 2515    2
> da <- read.csv("d-vix0411.csv",header=T) % Load csv data with names.
> head(da)  % See the first 6 rows
      Date VIX.Open VIX.High VIX.Low VIX.Close
1 1/2/2004    17.96    18.68   17.54     18.22
2 1/5/2004    18.45    18.49   17.44     17.49
 ....
6 1/9/2004    16.15    16.88   15.57     16.75
```

1.5 金融数据的例子

在本节中，我们研究金融时间序列收益率的变化，图 1-6 给出了从 2007 年 1 月 4 日至 2011 年 12 月 2 日苹果公司股票的日对数收益率的时序图. 根据前面的定义，日对数收益率简化为股票价格对数的变化. 在 R 中，序列的变化可以很容易通过对数价格的差(difference)得到. 具体来说，$r_t=\ln(p_t)-\ln(p_{t-1})$，这里 p_t 表示 t 时刻的股票价格. 在演示中，我们使用调整后的价格来计算股票的对数收益率，因为在样本期间，调整后的股票价格考虑了股票的分割. 从图 1-6 中可以看到：1)存在一些大的异常值；2)这些收益率序列在某些时期波动很大，而在其他时候是稳定的. 前者称为资产收益率存在着厚尾现象，后者称为收益率**波动集聚**(volatility clustering)现象.

图 1-7 显示了从 2007 年 1 月 4 日至 2011 年 12 月 2 日 10 年期国债到期收益率日变化的时序图，10 年期国债的到期收益率显示出和苹果公司股票日收益率相类似的特征. 图 1-8 提供了欧元/美元汇率日对数收益率的时序图. 同样，汇率的日对数收益率和股票的日对数收益率表现出了相同的特性. 图 1-9 中给出了欧元/美元的日汇率，数据来自 FRED.

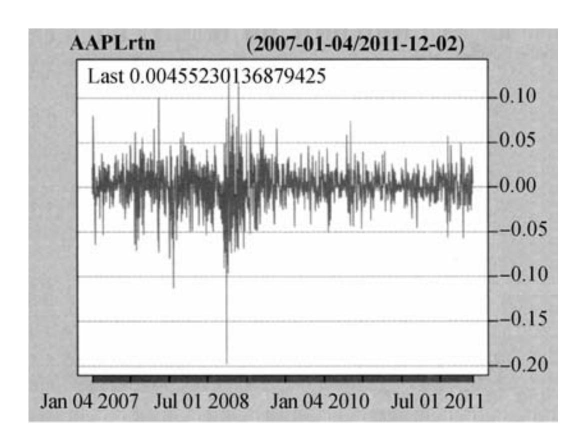


图 1-6 从 2007 年 1 月 4 日至 2011 年 12 月 2 日苹果公司股票的日对数收益率的时序图

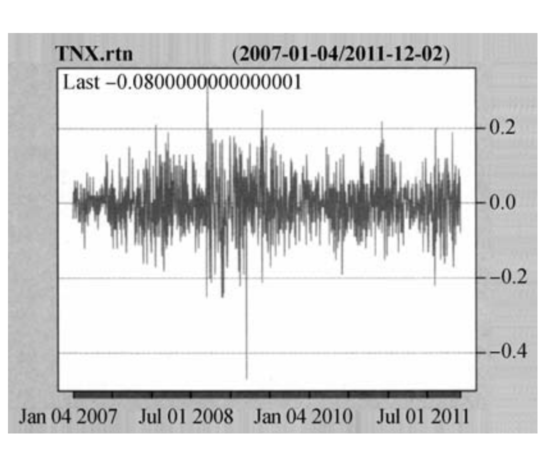


图 1-7 从 2007 年 1 月 4 日至 2011 年 12 月 2 日 10 年期国债的收益率日变化的时序图

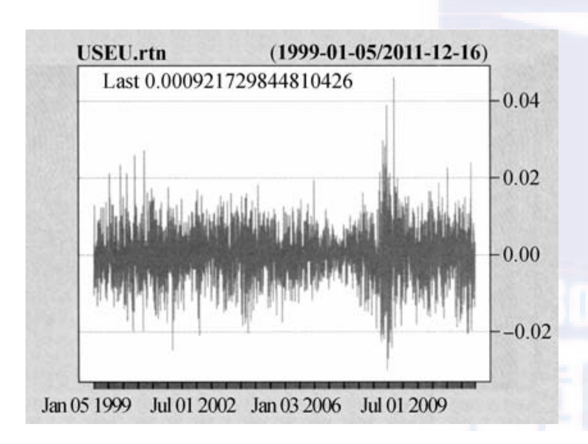


图 1-8 欧元/美元汇率日对数收益率的时序图

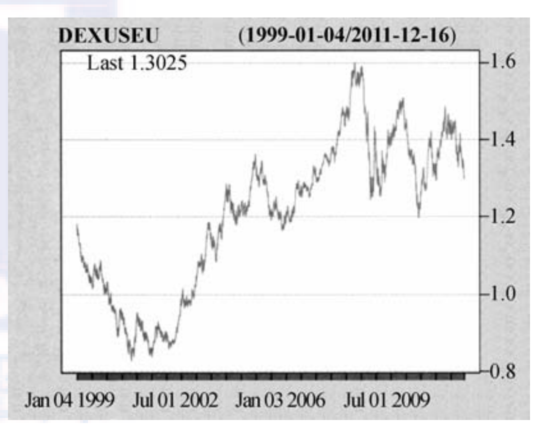


图 1-9 1999 年 1 月 4 日到 2011 年 12 月 16 日欧元/美元的日汇率

R 代码演示

```
> library(quantmod)
> getSymbols("AAPL",from="2007-01-03",to="2011-12-02") %Specify period
[1] "AAPL"
> AAPL.rtn=diff(log(AAPL$AAPL.Adjusted)) % Compute log returns
> chartSeries(AAPL.rtn,theme="white")
> getSymbols("^TNX",from="2007-01-03",to="2011-12-02")
[1] "TNX"

> TNX.rtn=diff(TNX$TNX.Adjusted) % Compute changes
> chartSeries(TNX.rtn,theme="white")
> getSymbols("DEXUSEU",src="FRED") % Obtain exchange rates from FRED
[1] "DEXUSEU"
> head(DEXUSEU)
           DEXUSEU
1999-01-04  1.1812
```

```
1999-01-05  1.1760
 ....
1999-01-11  1.1534
> tail(DEXUSEU)
           DEXUSEU
2011-12-09  1.3368
 ....
2011-12-16  1.3025
> USEU.rtn=diff(log(DEXUSEU$DEXUSEU))
> chartSeries(DEXUSEU,theme="white")
> chartSeries(USEU.rtn,theme="white")
```

1.6 收益率的分布性质

为了更好地理解资产收益率，我们先来研究这些资产收益率分布的性质．为了研究不同资产、不同时期收益率的行为表现，考虑一个 N 个资产的集合，在 T 个时间周期内持有这 N 个资产，即 $t=1,\cdots,T$．对每个资产 i，r_{it} 表示在 t 时刻的对数收益率．要研究的对数收益率为 $\{r_{it};\ i=1,\cdots,N;\ t=1,\cdots,T\}$．也可以考虑简单收益率 $\{R_{it};\ i=1,\cdots,N;\ t=1,\cdots,T\}$ 和对数超额收益率 $\{z_{it};\ i=1,\cdots,N;\ t=1,\cdots,T\}$．

统计分布及其矩的回顾

我们简要地回顾一下统计分布的基本性质和随机变量的矩．R^k 表示 k 维欧几里得空间，$\boldsymbol{x}\in R^k$ 表示 $\boldsymbol{x}$ 是 R^k 中的一个点．考虑两个随机向量 $\boldsymbol{X}=(X_1,\cdots,X_k)'$ 和 $\boldsymbol{Y}=(Y_1,\cdots,Y_q)'$．令 $P(\boldsymbol{X}\in A,\boldsymbol{Y}\in B)$ 表示 $\boldsymbol{X}$ 在子空间 $A\subset R^k$ 且 $\boldsymbol{Y}$ 在子空间 $B\subset R^q$ 中的概率．大多数时候，这里都假定随机向量是连续的．

联合分布

函数 $F_{X,Y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})=P(\boldsymbol{X}\leqslant\boldsymbol{x},\boldsymbol{Y}\leqslant\boldsymbol{y};\boldsymbol{\theta})$ 表示变量 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的参数为 $\boldsymbol{\theta}$ 的联合分布函数，其中 $\boldsymbol{x}\in R^p$，$\boldsymbol{y}\in R^q$，不等号"$\leqslant$"是分量对分量的运算．$\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的规律可由联合分布函数 $F_{X,Y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})$ 来刻画．例如，X 和 Y 之间的线性相依性可以由联合分布的协方差来给出．如果 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的联合概率密度函数 $f_{X,Y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})$ 存在，则

13~20

$$F_{X,Y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})=\int_{-\infty}^{\boldsymbol{x}}\int_{-\infty}^{\boldsymbol{y}}f_{x,y}(\boldsymbol{w},\boldsymbol{z};\boldsymbol{\theta})\mathrm{d}z\mathrm{d}w$$

这里，$\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 是连续随机向量．

边际分布

$\boldsymbol{X}$ 的边际分布为

$$F_X(\boldsymbol{x};\boldsymbol{\theta})=F_{X,Y}(\boldsymbol{x},\infty,\cdots,\infty;\boldsymbol{\theta})$$

这样，$\boldsymbol{X}$ 的边际分布可通过对 $\boldsymbol{Y}$ 求积分得到．$\boldsymbol{Y}$ 的边际分布也可以类似得到．

如果 $k=1$，X 是一个标量随机变量，其分布函数变成

$$F_X(x)=P(X\leqslant x;\boldsymbol{\theta})$$

称为 X 的**累积分布函数**（Cumulative Distribution Function，CDF），简称为分布函数．一个随机变量的分布函数是非递减的，即对于 $x_1\leqslant x_2$ 则有 $F_X(x_1)\leqslant F_X(x_2)$，且满足

$F_X(-\infty)=0$，$F_X(\infty)=1$. 对于给定的概率 p，使 $p\leqslant F_X(x_p)$成立的最小实数 x_p 称为随机变量 X 的第 p **分位数**(quantile)，更具体地，有

$$x_p = \inf_x\{x\,|\,p\leqslant F_X(x)\}$$

本书中我们用 CDF 来计算检验统计量的 p 值.

条件分布

给定 $\boldsymbol{Y}\leqslant\boldsymbol{y}$ 的条件下，$\boldsymbol{X}$ 的条件分布为

$$F_{X|Y\leqslant y}(\boldsymbol{x};\boldsymbol{\theta}) = \frac{P(\boldsymbol{X}\leqslant\boldsymbol{x},\boldsymbol{Y}\leqslant\boldsymbol{y};\boldsymbol{\theta})}{P(\boldsymbol{Y}\leqslant\boldsymbol{y};\boldsymbol{\theta})}$$

若概率密度函数存在，则在给定 $\boldsymbol{Y}=\boldsymbol{y}$ 的条件下，$\boldsymbol{X}$ 的条件密度为

$$f_{x|y}(\boldsymbol{x};\boldsymbol{\theta}) = \frac{f_{x,y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})}{f_y(\boldsymbol{y};\boldsymbol{\theta})} \tag{1-8}$$

其中，边际密度函数 $f_y(\boldsymbol{y};\boldsymbol{\theta})$可由下式得到

$$f_y(\boldsymbol{y};\boldsymbol{\theta}) = \int_{-\infty}^{\infty} f_{x,y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})\mathrm{d}\boldsymbol{x}$$

由式(1-8)可知，联合分布、边际分布和条件分布之间的关系可以表示为

$$f_{x,y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta}) = f_{x|y}(\boldsymbol{x};\boldsymbol{\theta})\times f_y(\boldsymbol{y};\boldsymbol{\theta}) \tag{1-9}$$

这个等式关系在时间序列分析中经常用到(如在极大似然估计时). 最后，$\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 是相互独立的随机向量，当且仅当 $f_{x|y}(\boldsymbol{x};\boldsymbol{\theta})=f_x(\boldsymbol{x};\boldsymbol{\theta})$. 这时，$f_{x,y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})=f_x(\boldsymbol{x};\boldsymbol{\theta})f_y(\boldsymbol{y};\boldsymbol{\theta})$.

21

随机变量的矩

一个连续型随机变量 X 的 ℓ 阶矩定义为

$$m'_\ell = E(X^\ell) = \int_{-\infty}^{\infty} x^\ell f(x)\mathrm{d}x$$

其中 E 表示期望(expectation)，$f(x)$是 X 的概率密度函数. 一阶矩称为 X 的**均值**(mean)或**期望**(expectation)，它用来度量该分布的中心位置，记为 μ_x. 对于资产，感兴趣的是它的均值是否为零. 换句话说，我们通常认为原假设为 $H_0:\mu_x=0$，其备择假设为 $H_a:\mu_x\neq 0$，或者 $H_0:\mu_x\leqslant 0$，其备择假设为 $H_a:\mu_x>0$.

第 ℓ 阶中心矩定义为

$$m_\ell = E[(X-\mu_x)^\ell] = \int_{-\infty}^{\infty}(x-\mu_x)^\ell f(x)\mathrm{d}x$$

这里假定上式中积分存在. 二阶中心矩可以度量 X 偏离其均值的程度，称为 X 的*方差*(variance)，记为 σ_x^2. 方差的正平方根 σ_x，称为 X 的*标准差*(standard deviation). 对于资产收益率，方差(或标准差)是测量不确定性的指标，因此经常用于资产的风险管理和度量. 随机变量的正态分布可由它的前两阶矩来确定. 而对于其他分布，可能需要更高阶的矩.

X 的三阶中心矩度量关于其均值的对称性，而四阶中心矩度量 X 的尾部特征. 在统计学中，标准化的三阶中心矩称为*偏度*(skewness)，标准化的四阶中心矩称为*峰度*(kurtosis)，它们分别用来描述随机变量的非对称程度和其尾部厚度. 具体地说，X 的偏度和峰度分别定义为

$$S(x)=E\left[\frac{(X-\mu_x)^3}{\sigma_x^3}\right],\quad K(x)=E\left[\frac{(X-\mu_x)^4}{\sigma_x^4}\right]$$

数量 $K(x)-3$ 称为**超额峰度**(excess kurtosis)，因为正态分布的峰度 $K(x)=3$. 这样，正态随机变量的超额峰度为 0. 若一个分布具有正的超额峰度，称为具有厚尾性. 也就是说，厚尾是指该分布在其尾部比正态分布有更多的“异常值”. 在实际中，这意味着来自这样分布的随机样本包含更多的极值，这样的分布称为尖峰的(leptokurtic). 另一方面，一个具有负超额峰度的分布是薄尾的(例如，有限区间上的均匀分布)，这样的分布称为**低峰的**(platykurtic). 在金融市场上，可以用随机变量的第一到第四阶矩来描述资产收益率的表现. 这并不意味着高阶矩不重要，只是它们学习起来更加困难.

在应用中，随机变量的矩可以用相应的样本偏度和样本峰度来估计. 设 $\{x_1, \cdots, x_T\}$ 是 X 的 T 个观测值，其样本均值表示为

22

$$\hat{\mu}_x=\frac{1}{T}\sum_{t=1}^{T}x_t \tag{1-10}$$

样本方差表示为

$$\hat{\sigma}_x^2=\frac{1}{T-1}\sum_{t=1}^{T}(x_t-\hat{\mu}_x)^2 \tag{1-11}$$

样本偏度表示为

$$\hat{S}(x)=\frac{1}{(T-1)\hat{\sigma}_x^3}\sum_{t=1}^{T}(x_t-\hat{\mu}_x)^3 \tag{1-12}$$

样本峰度表示为

$$\hat{K}(x)=\frac{1}{(T-1)\hat{\sigma}_x^4}\sum_{t=1}^{T}(x_t-\hat{\mu}_x)^4 \tag{1-13}$$

在相当弱的条件下，样本均值 $\hat{\mu}_x$ 是 μ_x 的一致估计，表示当 $T\to\infty$ 时，$\hat{\mu}_x$ 收敛到 μ_x. 更具体地说，当 T 充分大时，我们有 $\hat{\mu}_x\sim N(\mu_x, \sigma_x^2/T)$，该结果经常用于有关 μ_x 的假设检验. 例如，考虑原假设 $H_0: \mu_x=0$，其备择假设为 $H_a: \mu_x\neq 0$. 检验统计量为

$$t=\frac{\sqrt{T}\,\hat{\mu}_x}{\hat{\sigma}_x}$$

服从一个自由度为 $T-1$ 的学生 t 分布. 对于充分大的时间 T，其检验统计量接近一个标准正态分布. 在显著性水平为 $100\alpha\%$ 时，如果满足 $|t|>Z_{1-\alpha/2}$，那么该条规则可以用来拒绝原假设 H_0，其中 $Z_{1-\alpha/2}$ 是标准正态分布的第 $(1-\alpha/2)$ 分位数. 大多数统计软件包可以提供每个检验统计量的伴随概率的 p 值. 在显著性水平为 $100\alpha\%$ 时，如果检验统计量的 p 值小于显著性水平 α，那么决策规则为拒绝原假设 H_0.

如果 X 是一个正态随机变量，则 $\hat{S}(x)$ 和 $\hat{K}(x)-3$ 的分布渐近为均值为零、方差分别是 $6/T$ 和 $24/T$ 的正态分布，具体参见 Snedecor 和 Cochran(1980，第 78 页). 我们可以用这些渐近分布的性质来检验资产收益率是否具有正态性. 给定一个资产收益率序列 $\{r_1, \cdots, r_T\}$，要检验其偏度，考虑原假设 $H_0: S(r)=0$ 和备择假设 $H_a: S(r)\neq 0$. 由式(1-12)定义的样本偏度的 t 比为

23

$$t=\frac{\hat{S}(r)}{\sqrt{6/T}}$$

决策规则如下：在显著性水平为 α 时，若满足 $|t|>Z_{1-\alpha/2}$，则拒绝原假设.

类似地，我们可以检验收益率序列的超额峰度，假设 $H_0: K(r)-3=0$，其备择假设为 $H_a: K(r)-3\neq 0$. 则检验统计量为

$$t=\frac{\hat{K}(r)-3}{\sqrt{24/T}}$$

该统计量服从一个渐近的标准正态分布. 当且仅当检验统计量的 p 值小于显著性水平 α 时，决策规则拒绝原假设 H_0. Jarque 和 Bera(1987)将两个先验检验结合在一起检验 r_t 的正态性，则检验统计量为

$$JB=\frac{\hat{S}^2(r)}{6/T}+\frac{(\hat{K}(r)-3)^2}{24/T}$$

该统计量渐近为一个自由度为 2 的 χ^2 随机变量. 当 JB 统计量的 p 值小于显著性水平 α 时，则拒绝正态性分布的原假设.

例 1.2 考虑 3M 公司股票从 2001 年 1 月 2 日至 2011 年 9 月 30 日的日简单收益率的数据. 该数据来自芝加哥大学证券价格研究中心(CRSP). 图 1-10 给出了数据的时序图. 这里我们使用 Rmetrics 中的 R 添加包 fBasics 的 `basicStats` 命令，计算收益率的汇总统计数据，并进行一些基本假设检验. 从上述结果可知，共有 2704 个数据点，简单收益率的样本均值为 0.0278%，样本标准差为 0.0155. 样本偏度和超额峰度分别为 0.0279 和 4.631. 下面考虑原假设 $H_0: \mu=0$，其备择假设为 $H_a: \mu\neq 0$，其中 μ 表示 3M 公司股票日简单收益率的平均值. 其检验统计量为

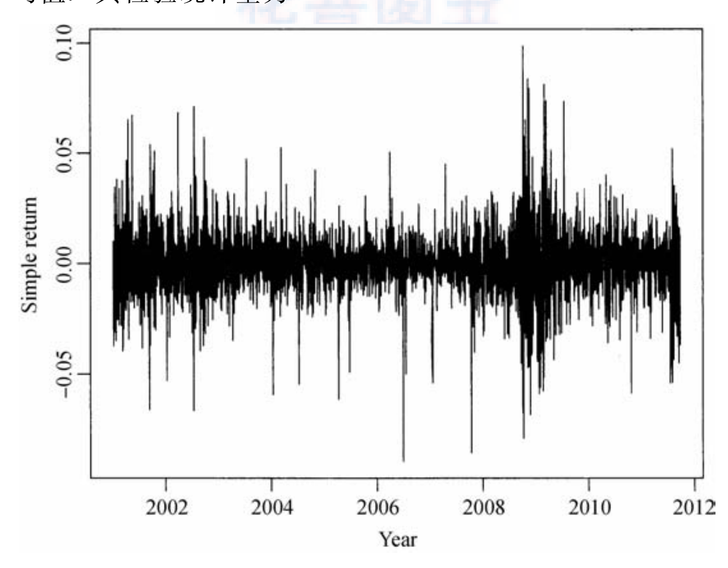


图 1-10 从 2001 年 1 月 2 日至 2011 年 9 月 30 日 3M 公司股票日简单收益率数据的时序图

$$t=\frac{0.000\,278}{0.0155/\sqrt{2704}}=0.933$$

这里 p 值为 0.35，大于 0.05．因此，在 5%的显著性水平下，我们不能拒绝原假设．对于偏度，原假设为 H_0：$S=0$，其备择假设为 H_a：$S\neq0$．其检验统计量为

$$t=\frac{0.0279}{\sqrt{6/2704}}=0.59$$

p 值为 0.55．而且，在 5%的显著性水平下，我们不能拒绝原假设．对于超额峰度，原假设为 H_0：$K-3=0$，其备择假设为 H_a：$K-3\neq0$．对于 3M 公司的股票收益率，其检验统计量为

$$t=\frac{4.631}{\sqrt{24/2704}}=49.15$$

比标准正态分布大，其 p 值接近于零，因此，可以拒绝 $K=3$ 的原假设．换言之，3M 公司股票的日简单收益有一个较大的尾部．最后，Jarque-Bera 检验统计量是 2422，比自由度为 2 的 χ^2 分布大．因此，3M 公司日简单收益的正态性假设被拒绝．该收益率有较大的厚尾现象并不奇怪． ■

R 代码演示（输出经过编辑）

```
> library(fBasics) % Load package
> da=read.table("d-mmm-0111.txt",header=T) % Load data


> head(da) % Show the first 6 rows of data
      date        rtn
1 20010102 -0.010892
  ....
6 20010109 -0.015727
> mmm=da[,2]  % Obtain 3m simple returns
> basicStats(mmm) %Compute summary statistics
                  mmm
nobs          2704.000000 % Sample size
NAs              0.000000 % No of missing values
Minimum         -0.089569 % Minimum
Maximum          0.098784 % Maximum
1. Quartile     -0.007161 % 25th percentile
3. Quartile      0.007987 % 75th percentile
Mean             0.000278 % Sample mean
Median           0.000350 % Sample median
Sum              0.751082 % Sample total
SE Mean          0.000298 % Standard error of Sample mean
                          % = sqrt(sample variance/sample size]
LCL Mean        -0.000306 % Lower bound of 95% C.I.
UCL Mean         0.000862 % Upper bound of 95% C.I.
Variance         0.000240 % Sample variance
Stdev            0.015488 % Sample standard error
Skewness         0.027949 % Sample skewness
Kurtosis         4.630925 & % Sample excess kurtosis

% Commands for individual moments
> mean(mmm)
[1] 0.000277767
```

```
> var(mmm)
[1] 0.0002398835
> stdev(mmm) % standard deviation
[1] 0.01548817
% Simple tests
> t.test(mmm)  % Testing mean return = 0
        One Sample t-test
data:  mmm
t = 0.9326, df = 2703, p-value = 0.3511
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.0003062688  0.0008618028 % See prior summary statistics.
% p-value > 0.05; one cannot reject the null hypothesis.

> s3=skewness(mmm)
> T=length(mmm) % Sample size
> T
[1] 2704
> t3=s3/sqrt(6/T) % Skewness test
> t3
[1] 0.593333
> pp=2*(1-pnorm(t3)) % Compute p-value
> pp
[1] 0.5529583 % Cannot reject the null of symmetry.
> s4=kurtosis(mmm)
> t4=s4/sqrt(24/T) % Kurtosis test
> t4
[1] 49.15475 % Value is huge; reject the null. Has heavy tails.

> normalTest(mmm,method='jb') % JB-test
Title:  Jarque - Bera Normalality Test
Test Results:
  STATISTIC:   X-squared: 2422.4384
  P VALUE:  Asymptotic p Value: < 2.2e-16 % Reject normality
```


1.7 金融数据的可视化

在分析金融数据时，图示法是一个有用的工具．除了前面显示的时间序列图外，本节将讨论其他的金融数据制图方法．为了获得更好的资产收益分布率的可视化方法，我们检验这些数据的直方图或经验密度函数．例如，从 2001 年 1 月 2 日至 2011 年 9 月 30 日 3M 公司股票的日简单收益率，共 2704 个观测值．前面已给出汇总的统计数据．图 1-11 显示数据的直方图，它通过将数据划分为 30 个子区间而得到．由图 1-11 可知，收益率曲线关于它的零均值对称．图 1-12 中的实线列出了 3M 公司股票收益率的经验密度函数，通过用非参数的平滑方法得到．这些经验密度函数可以视为直方图的一个精确翻版．图 1-12 中的虚线显示了 3M 公司股票正态分布的密度函数，它们具有相同的均值和标准差，这些时序图提供了 3M 公司股票简单收益率正态假设的可视化过程．和正态分布相比，经验密度函数具有更高的峰值和更厚的尾部．这种现象对于股票收益的日数据来说很常见．一般情况下，实线和虚线之间存在偏差，说明 3M 公司股票的日简单收益率是非正态的．这和我们前面正态性检验的结果相一致．

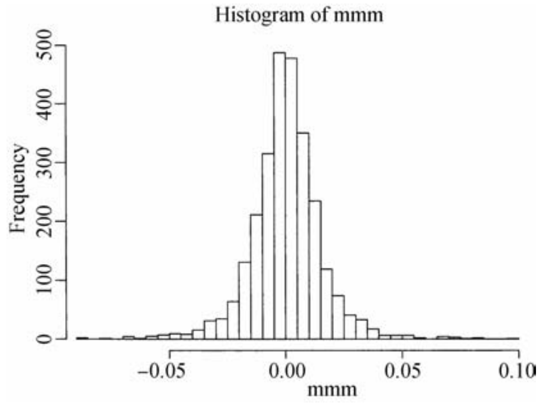


图 1-11　从 2001 年 1 月 2 日至 2011 年 9 月 30 日 3M 公司股票的日简单收益率的直方图

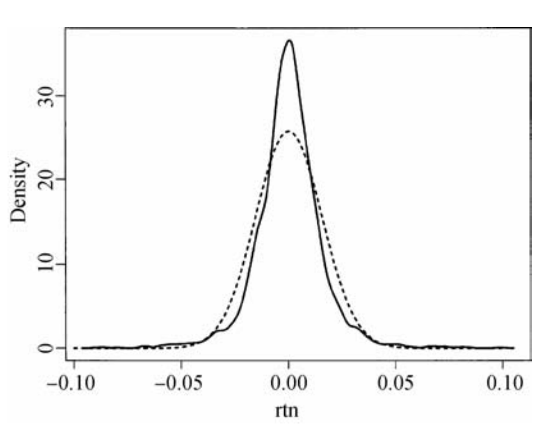


图 1-12　从 2001 年 1 月 2 日至 2011 年 9 月 30 日 3M 公司股票的日简单收益率的经验密度函数，虚线是正态分布的密度函数，它们具有相同的均值和标准差

为了研究股票的价格波动，我们考虑股票的日开盘价、最高价、最低价和收盘价. 图 1-13 显示苹果公司从 2011 年 1 月 3 至 6 月 30 日股票数据的时序图. 在文献中该图称为条形图(Bar Chart). 我们使用 R 脚本 `ohlc.R` 画出该图. 该脚本的修改版由 Klemelä (2009)完成. 在图 1-13 中，竖线表示日股票的价格范围，竖线左侧水平线上的点给出了股票的开盘价，右边给出了收盘价. 在时序图中，该曲线提供的信息有限，无法给出太多的天数. 图 1-14 显示从 2010 年 1 月 2 日至 2011 年 12 月 8 日苹果公司股票的日收盘价，同时也给出了过去 21 个交易日收盘价格的移动平均，称为**移动平均曲线**(moving-average chart). 这里使用 21 天是任选的，在一个月中的交易日是个粗略的数字. 移动平均曲线提供了相对最近的历史股价信息. 在统计上，移动平均是减少随机波动的一个简单统计方法.

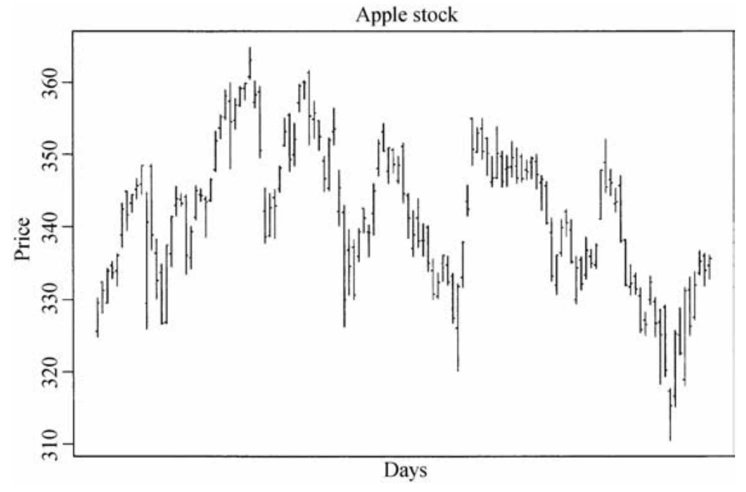


图 1-13　苹果公司从 2011 年 1 月 3 至 6 月 30 日股票数据条形图，竖线表示日股票的价格范围，竖线左侧水平线上的点给出了股票的开盘价，右边给出了收盘价

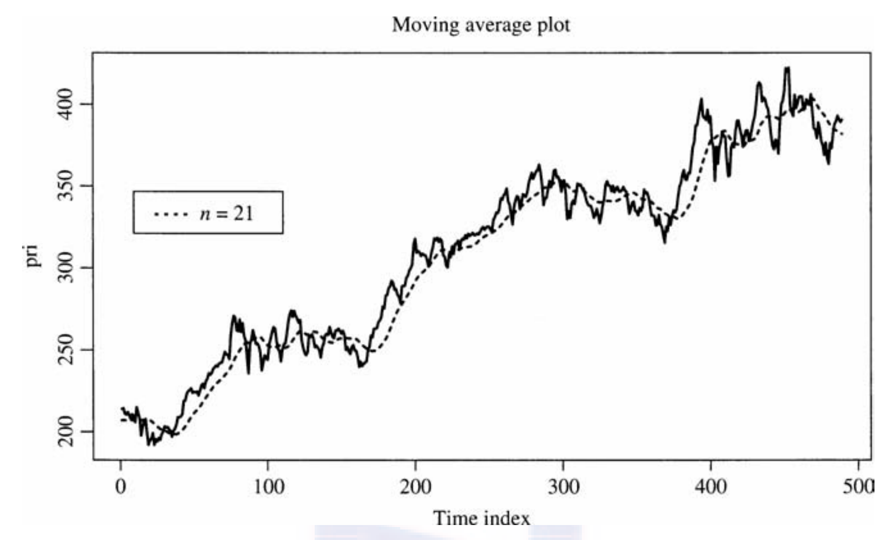


图 1-14　从 2010 年 1 月 2 日至 2011 年 12 月 8 日苹果公司股票日收盘价，在过去 21 个交易日的移动平均曲线

对于多个资产收益率，图 1-15 显示了从 1926 年 1 月至 2011 年 9 月，IBM 公司股票和 S&P 综合指数月对数收益率的时序图．数据来自美国股市资料库(CRSP)．除了经济**大萧条**(Great Depression)时期外，单个股票的收益率一般比市场指数存在更大的波动．在时序图中 IBM 公司的股票和市场指数收益率之间存在一定的下降或跳跃．图 1-16 显示了两个对数收益率的**散点图**(scatter plot)．散点图还显示了两个收益率的最小二乘线性回归．正如所料，IBM 股票和市场指数收益率之间有正的相关关系．这种线性关系可以度量两个收益率之间的相关性．在这种情况下，其相关系数为 0.64．同时，我们也可以考虑下面的**市**

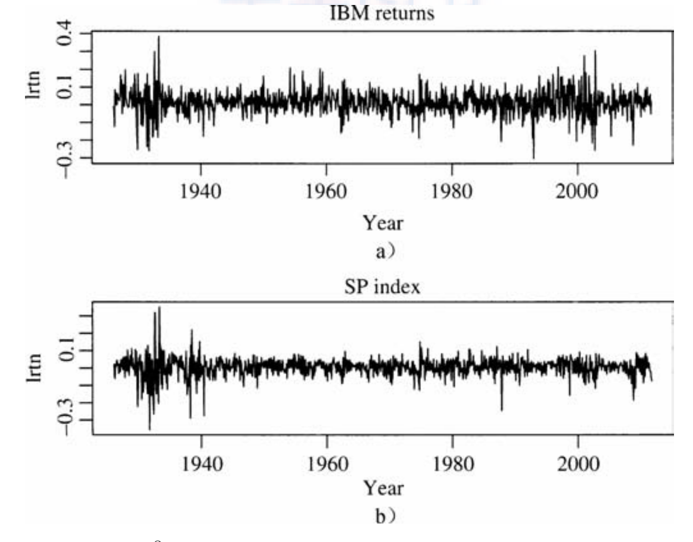


图 1-15　IBM 股票和标普(S&P)综合指数从 1926 年 1 月至 2011 年 9 月的月对数收益率时序图

场模型(Market Model)

$$r_t = \alpha + \beta m_t + \varepsilon_t$$

这里，r_t 和 m_t 分别表示单个股票的资产收益率和市场收益率，ε_t 是误差项．参数 α 表示相对于市场收益的超额收益，β 为系数．对于 IBM 股票月对数收益率，我们有 $r_t = 0.008 + 0.807m_t + \varepsilon_t$．这两个参数在 5%的显著性水平下通常显著不为零．有关市场模型的更多信息，请参阅 Sharpe(1964)的资本资产定价模型(Capital Asset Pricing Model，CAPM)．

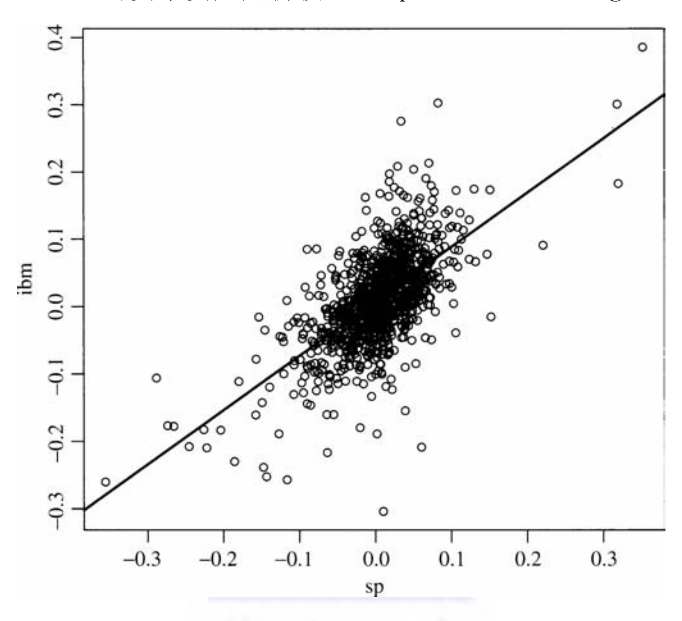


图 1-16　从 1926 年 1 月至 2011 年 9 月 IBM 公司股票和 S&P 综合指数月对数收益率的散点图，实线是两个收益率最小二乘线性拟合

R 代码演示

```
> library(fBasics)
> da=read.table("d-mmm-0111.txt",header=T) % Load data
> mmm=da[,2] % Locate 3M simple returns
> hist(mmm,nclass=30) % Histogram
> d1=density(mmm)  % Obtain density estimate
> range(mmm)  % Range of 3M returns
[1] -0.089569  0.098784
> x=seq(-.1,.1,.001) % Create a sequence of x with increment 0.001.
     % The next command creates normal density
> y1=dnorm(x,mean(mmm),stdev(mmm))
> plot(d1$x,d1$y,xlab='rtn',ylab='density',type='l')
> lines(x,y1,lty=2)
 % ohlc plot
> library(quantmod)
> getSymbols("AAPL",from="2011-01-03",to="2011-06-30")
> X=AAPL[,1:4] % Locate open, high, low, and close prices
> xx=cbind(as.numeric(X[,1]),as.numeric(X[,2]),as.numeric(X[,3]),
        as.numeric(X[,4]))
> source("ohlc.R") % Compile the R script
```

```
> ohlc(xx,xl="days",yl="price",title="Apple Stock")
 % Moving average plot
> source("ma.R")  % Compile R script
> getSymbols("AAPL",from="2010-01-02",to="2011-12-08")
> x1=as.numeric(AAPL$AAPL.Close) % Locate close price
> ma(x1,21)
 % Bivariate and Scatter plots
> da=read.table("m-ibmsp-2611.txt",header=T)
> head(da)
      data       ibm        sp
1 19260130 -0.010381  0.022472
 .....
6 19260630  0.068493  0.043184
> ibm=log(da$ibm+1) % Transform to log returns
> sp=log(da$sp+1)
> tdx=c(1:nrow(da))/12+1926 % Create time index
> par(mfcol=c(2,1))
> plot(tdx,ibm,xlab='year',ylab='lrtn',type='l')
> title(main='(a) IBM returns')
> plot(tdx,sp,xlab='year',ylab='lrtn',type='l') % X-axis first.
> title(main='(b) SP index')
> cor(ibm,sp)  % Obtain sample correlation
[1] 0.6409642
> m1=lm(ibm~ sp)  % Fit the Market Model (linear model)
> summary(m1)
Call: lm(formula = ibm ~ sp)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.007768   0.001672   4.645 3.84e-06 ***
sp          0.806685   0.030144  26.761 < 2e-16  ***
---
Residual standard error: 0.05348 on 1027 degrees of freedom
Multiple R-squared: 0.4108,     Adjusted R-squared: 0.4103
> plot(sp,ibm,cex=0.8)  % Obtain scatter plot
> abline(0.008,.807) % Add the linear regression line
```

1.8 一些统计分布

在文献中提到了一些关于资产收益率边际分布的统计分布，包括正态分布、对数正态分布、稳态分布、正态分布的尺度混合．下面我们简要地介绍这些分布．

1.8.1 正态分布

在金融研究中，传统的假设是：简单收益率$\{R_{it}|t=1,\cdots,T\}$是独立同分布(independently and identically distributed，iid)，且都服从一个固定均值和方差的正态分布．这个假设使得资产收益率的统计性质变得易于处理，同时也遇到了几个困难：第一，简单资产收益率的下界为-1，而正态分布可以取实轴上的任意值，没有下界；第二，如果R_{it}是正态分布，那么多期的简单收益率$R_{it}[k]$是单期收益率的乘积，不再是正态分布；第三，很多资产收益率都存在正的超额峰度，因为很多实证结果均不支持正态性的假设．

1.8.2 对数正态分布

另一个常用的假定是：资产的对数收益率r_t是独立同分布，且服从一个均值为μ、方

24 ~ 32

差为 σ^2 的正态分布．简单收益率是独立同分布的对数正态随机变量，其均值和方差分别为

$$E(R_t)=\exp\left(\mu+\frac{\sigma^2}{2}\right)-1,\quad \mathrm{Var}(R_t)=\exp(2\mu+\sigma^2)[\exp(\sigma^2)-1] \tag{1-14}$$

这两个等式在研究资产收益率时常用(如在用对数收益率建立的模型进行预测时)．反之，假设简单收益率 R_t 服从对数正态分布，其均值为 m_1，方差为 m_2，则相应的对数收益率 r_t 的均值和方差分别为

$$E(r_t)=\ln\left(\frac{m_1+1}{\sqrt{1+\frac{m_2}{(1+m_1)^2}}}\right),\quad \mathrm{Var}(r_t)=\ln\left(1+\frac{m_2}{(1+m_1)^2}\right)$$

因为有限个独立同分布的正态随机变量之和仍服从正态分布，所以在 $\{r_t\}$ 是正态分布假设下，$r_t[k]$ 也服从正态分布．另外，r_t 没有下界，由 $1+R_t=\exp(r_t)$，可得 R_t 的下界．然而，对数正态分布的假设与历史股票收益率的所有性质并不都相符，很多股票收益率表现出具有正的超额峰度．

1.8.3 稳态分布

稳态分布是正态分布的自然推广，它们在加法运算下是稳定的，这一点满足连续复合收益率 r_t 的要求．而且，稳态分布能刻画股票的历史收益率所显现出来的超额峰度．然而，非正态的稳态分布没有有限方差，这一点与大部分金融理论相矛盾．另外，用非正态的稳态分布进行统计建模是很困难的．非正态稳态分布的例子是柯西分布(Cauchy distribution)，它关于自己的中位数对称，但其方差是无限的．

1.8.4 正态分布的尺度混合

在最近股票收益率研究中，人们倾向于利用正态分布的尺度混合或有限混合．在正态分布的尺度混合假定下，对数收益率 r_t 服从均值为 μ、方差为 σ^2 的正态分布(即 $r_t\sim N(\mu,\sigma^2)$)．但是，σ^2 是一个随机变量，服从一个正的分布(如 σ^{-2} 服从一个 Gamma 分布)．正态分布的一个有限混合的例子为

$$r_t\sim(1-X)N(\mu,\sigma_1^2)+XN(\mu,\sigma_2^2)$$

其中，X 是一个服从 Bernoulli 分布的随机变量，即 $P(X=1)=\alpha$，$P(X=0)=1-\alpha$，且 $0<\alpha<1$，这里 σ_1^2 较小，而 σ_2^2 相对较大．例如，对 $\alpha=0.05$，有限混合分布指出 95% 的收益率服从 $N(\mu,\sigma_1^2)$，还有 5% 的收益率服从 $N(\mu,\sigma_2^2)$．σ_2^2 的较大值能使混合分布把更多的异常值放在其分布的尾部．来自 $N(\mu,\sigma_2^2)$ 的收益率的百分比较低，说明大多数收益率服从一个简单的正态分布．混合正态分布的优点是它们保持了正态分布的易处理性、具有有限高阶矩且能刻画这些超额峰度．然而，我们很难估计其混合参数(如有限混合分布中的参数 α)．

33

图 1-17 显示的是有限混合正态分布、柯西分布和标准正态分布的概率密度函数．有限混合的正态分布为 $(1-X)N(0,1)+XN(0,16)$，其中 X 是一个 Bernoulli 随机变量，$P(X=1)=0.05$．柯西分布的密度函数是

$$f(x)=\frac{1}{\pi(1+x^2)}\quad -\infty<x<\infty$$

可见，柯西分布比有限混合的正态分布具有更厚的尾部，而有限混合的正态分布比标准正态分布存在更厚的尾部.

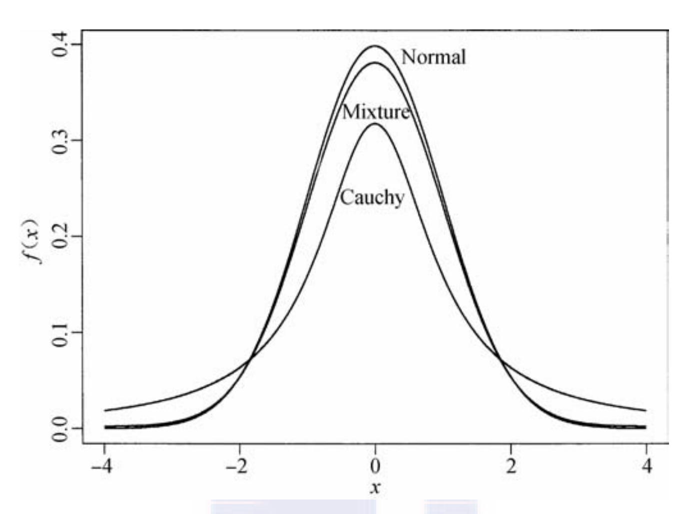


图 1-17　有限混合的正态分布、稳态分布和标准正态分布密度函数之间的比较

1.8.5　多元收益率

设 $\boldsymbol{r}_t=(r_{1t},\ \cdots,\ r_{Nt})'$是 N 个资产在 t 时刻的对数收益率. 多元分布关心$\{\boldsymbol{r}_t\}_{t=1}^{T}$的联合分布. 在序列相关性存在的情况下，统计分析主要集中在条件分布函数 $F(\boldsymbol{r}_t|\boldsymbol{r}_{t-1},\ \cdots,\ \boldsymbol{r}_1,\ \boldsymbol{\theta})$具体形式的设定上. 尤其是，在投资组合和风险管理中，人们最关心的是 $\boldsymbol{r}_t$ 的条件期望和条件协方差矩阵随时间怎样变化.

34

随机向量 $\boldsymbol{X}=(X_1,\ \cdots,\ X_p)$的均值向量和协方差矩阵定义为

$$E(\boldsymbol{x})=\boldsymbol{\mu}_x=[E(X_1),\cdots,E(X_p)]'$$

$$\mathrm{Cov}(\boldsymbol{x})=\boldsymbol{\Sigma}_x=E[(\boldsymbol{x}-\boldsymbol{\mu}_x)(\boldsymbol{x}-\boldsymbol{\mu}_x)']$$

假设 $\boldsymbol{X}$ 的期望存在. 当 $\boldsymbol{X}\in\{\boldsymbol{x}_1,\ \cdots,\ \boldsymbol{x}_T\}$时，样本均值和样本协方差矩阵定义为

$$\hat{\boldsymbol{\mu}}_x=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_t,\quad \hat{\boldsymbol{\Sigma}}_x=\frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{x}_t-\hat{\boldsymbol{\mu}}_x)(\boldsymbol{x}_t-\hat{\boldsymbol{\mu}}_x)'$$

假设 $\boldsymbol{X}$ 的协方差矩阵存在，那么这些样本统计量都是它们理论上的一致估计. 在金融文献中，多元正态分布常用其对数收益率 $\boldsymbol{r}_t$ 来表示.

为了证明这一点，下面考虑 IBM 股票和 S&P 500 综合指数的月对数收益率(从 1926 年 1 月至 2011 年 9 月)，如图 1-16 所示. 令 $\boldsymbol{r}_t=(r_{1t},\ r_{2t})'$，这里 r_{1t}和 r_{2t}分别是 IBM 股票和 S&P 500 综合指数的月对数收益率，共有 1029 个观测值. 则 $\boldsymbol{r}_t$ 的样本均值和协方差矩阵为

$$\hat{\boldsymbol{\mu}}=\begin{bmatrix}0.0113\\0.0044\end{bmatrix},\quad \hat{\boldsymbol{\Sigma}}=\begin{bmatrix}4849 & 2470\\2470 & 3062\end{bmatrix}\times 10^{-6}$$

要检查二元正态假设的有效性，可以使用统计模拟的方法. 从一个二元正态分布中，我们产生了 1029 个数据，其均值为$\hat{\boldsymbol{\mu}}$，协方差矩阵为$\hat{\boldsymbol{\Sigma}}$. 在 R 添加包中，可以使用

mnormt 中的 rmnorm 命令来实现. 图 1-18 显示了模拟过程的散点图. 通过比较图 1-18 和图 1-16 中的散点图，我们发现这两个时序图之间存在着显著的差异，这就表明这里的正态性假设是值得商榷的.

R 代码演示

```
> da=read.table("m-ibmsp-2611.txt",header=T) % Load  data
> dim(da)
[1] 1029    3
> ibm=log(da$ibm+1) % Compute log returns
> sp=log(da$sp+1)
> rt=cbind(ibm,sp) % Obtain bivariate returns
> m1=apply(rt,2,mean) % Obtain sample means
> v1=cov(rt) % Obtain sample covariance matrix
> m1
        ibm          sp
0.011303024 0.004381644
> v1
            ibm          sp
ibm 0.004849390 0.002469738
sp  0.002469738 0.003061590
> library(mnormt) % Load package
> x=rmnorm(1029,mean=m1,varcov=v1) % Simulation
> dim(x)
[1] 1029    2
> plot(x[,2],x[,1],xlab='sim-sp',ylab='sim-ibm',cex=0.8]
```

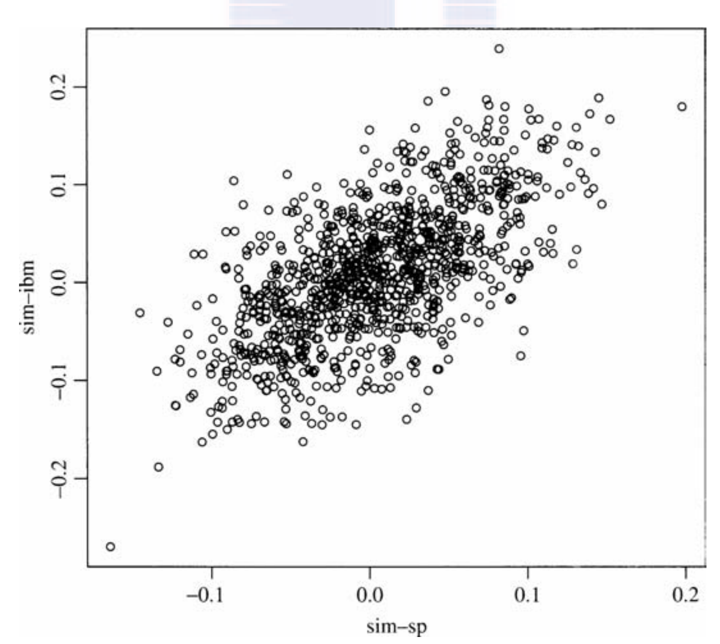


图 1-18 根据二元正态分布模拟产生的 1029 个数据的散点图，该二元正态分布基于 IBM 股票和标普(S&P)综合指数的月对数收益率的样本均值和协方差

习题

1. 考虑从 2001 年 9 月 1 日到 2011 年 9 月 30 日美国运通公司(AXP)、CRSP 价值权重指数(VW)、CRSP 的等权重指数(EW)以及 S&P 综合指数的日简单收益率. 收益率中包含有支付的股息. 数据来自 d-axp3dx-0111.txt(date, axp, vw, ew, sp).
   (a) 计算每个收益率序列的样本均值、标准差、偏度、超额峰度、最大值和最小值.
   (b) 把简单收益率转换成对数收益率. 计算每个对数收益率的样本均值、标准差、偏度、超额峰度、最大值和最小值.
   (c) 对 AXP 股票的对数收益的均值等于零的原假设进行检验. 在 5%的显著性水平下，得出你的结论.

35~36

2. 考虑从 1940 年 1 月到 2011 年 9 月通用电器(GE)、CRSP 价值权重指数(VW)、CRSP 等权重指数(EW)以及 S&P 综合指数的月收益率数据. 重新回答习题 1 中的所有问题，这里收益率中包含有股息，数据来自 m-ge3dx-4011.txt(date, ge, vw, ew, sp).
3. 考虑从 1940 年 1 月到 2011 年 9 月 S&P 综合指数的月股票收益率，进行下面的检验，在 5%的显著性水平下得出你的结论.
   (a) 检验假设 $H_0: \mu=0$，其备择假设为 $H_a: \mu\neq 0$，这里 μ 是收益率的均值.
   (b) 检验假设 $H_0: m_3=0$，其备择假设为 $H_a: m_3\neq 0$，这里 m_3 是收益率的偏度.
   (c) 检验假设 $H_0: K=3$，其备择假设为 $H_a: K\neq 3$，这里 K 是收益率的峰度.
4. 再次考虑习题 1 中美国运通公司从 2001 年 9 月 1 日到 2011 年 9 月 30 日股票的日对数收益率. 在 5% 的显著性水平下，进行下述假设检验：
   (a) 原假设是：收益率的偏度度量等于零.
   (b) 原假设是：收益的超额峰度等于零.
5. 从芝加哥的联邦储备银行得到日汇率，数据是经过纽约联邦储备银行认证的纽约市每日中午买入价. 考虑从 2007 年 1 月 2 日到 2011 年 11 月 30 日美元对英镑、日元的汇率，这些数据可以从网站下载. 计算：
   (a) 每个汇率的日对数收益率.
   (b) 汇率的日对数收益率的样本均值、标准差、偏度、超额峰度、最大值和最小值.
   (c) 画出美元/日元汇率的日对数收益率的密度函数.
   (d) 检验假设 $H_0: \mu=0$，其备择假设为 $H_a: \mu\neq 0$，这里 μ 表示美元/日元汇率的日对数收益率的均值. 在 5%的显著性水平下给出你的结论.

参考文献

Adler J. R in a Nutshell. Sebastopol (CA): O'Reilly Media; 2010.

Campbell JY, Lo AW, MacKinlay AC. The Econometrics of Financial Markets. Princeton (NJ): Princeton University Press; 1997.

Crawley MJ. The R Book. Hoboken (NJ): John Wiley & Sons; 2007.

Hull JC. Options, Futures, and Other Derivatives. 8th ed. Upper Saddle River (NJ): Prentice Hall; 2011.

Jarque CM, Bera AK. A test of normality of observations and regression residuals. Int Stat Rev 1987; 55:163172.

Klemelä J. Smoothing of Multivariate Data: Density Estimation and Visualization. Hoboken (NJ): John Wiley & Sons; 2009.

Sharpe W. Capital asset prices: a theory of market equilibrium under conditions of risk. J Finance 1964; 19:425–442.

Snedecor GW, Cochran WG. chapStatistical Methods. 7th ed. Ames (IA): Iowa State University Press; 1980.

37

第2章　金融时间序列的线性模型

在这一章中，我们讨论金融时间序列建模和预测中有用的方法和线性模型．我们应用实例来介绍一些重要的统计概念，并一步步地进行数据分析，同时讨论它们在金融中的应用．线性时间序列分析的基本概念，参见 Tsay(2010，第 2 章)、Box 等(1994，第 2 章和第 3 章)、Brockwell 和 Davis(2002，第 1～3 章)、Shumway 和 Stoffer(2000)以及 Woodward 等(2012)．

本章介绍的模型包括：1)简单自回归(AR)模型；2)简单移动平均(MA)模型；3)混合自回归移动平均(ARMA)模型；4)包含单位根检验的单位根模型；5)指数平滑模型；6)季节模型；7)带时间序列误差的回归模型；8)对长期相依性的分数阶差分模型．对每一个模型，我们学习它们的基本性质，包括模型选择方法、产生预测的方式并讨论它们的应用．本章也讨论了如何比较不同的模型，例如回测检验和模型平均预测．

设$\{x_t\}$是与时间有关的金融度量的集合．图 2-1 给出了从 2003 年 1 月 3 日到 2010 年 4 月 5 日期间苹果公司股票的日收盘价．图 2-1 显示日收盘价有某种程度的波动性，并在取样区间内有一个向上的运动趋势．图 2-2 给出了 1983 年到 2009 年可口可乐公司每股股票的季度盈利．图 2-2 中标出了具体的季度．除了有向上的趋势外，收益也同时显示有明显的年度模式，这在时间序列中称为季节性(seasonality)．后面将看到，许多经济和金融时间序列呈现明显的季节趋势．图 2-3 给出了标普 500(S&P 500)指数在 1926 年 1 月到 2009 年 12 月的月对数收益率．从图 2-3 中可以看出，收益率在 0 值周围波动，并且除了少数几个极端值以外，它们都在一个固定范围内波动．图 2-4 给出了两个时间序列，它们分别是美国 3 月期和 6 月期国债在 1959 年 1 月 2 日到 2010 年 4 月 16 日的周利率．这些利率取自二级市场．其中，图 2-4 的上图是 6 月期国债利率，下图是 3 月期国债利率．两个利率序列的移动模式

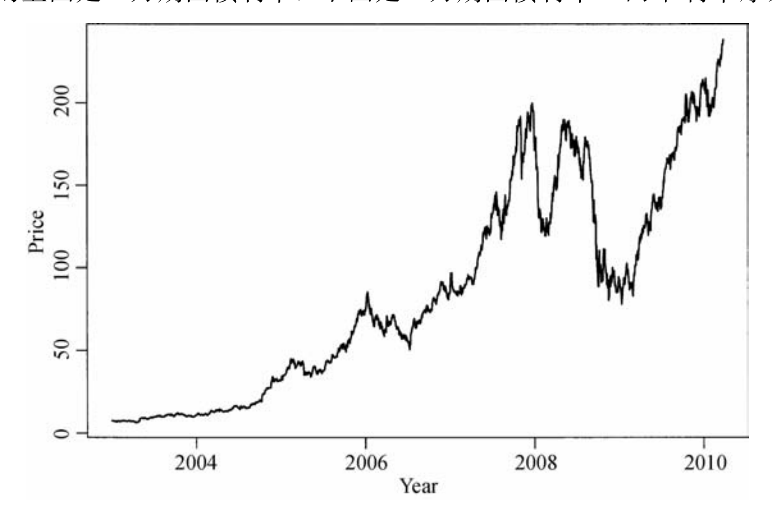


图 2-1　苹果公司股票在 2003 年 1 月 3 日到 2010 年 4 月 5 日期间的日收盘价

近似，但是也呈现出某种差异．如预期的那样，6 月期国债利率通常高于 3 月期，但是在某些时期上，3 月期的国债利率较高，例如 20 世纪 80 年代的早期．在上述 4 个例子中，序列 x_t 大致上是在相等分隔的时间区间上观测的．这种现象在利率期限结构中被称为**收益率曲线倒挂**．这 4 个序列就是本章要分析的金融时间序列的例子．我们的目的是学习序列的动态相依性，据此对序列做出适当的推断．

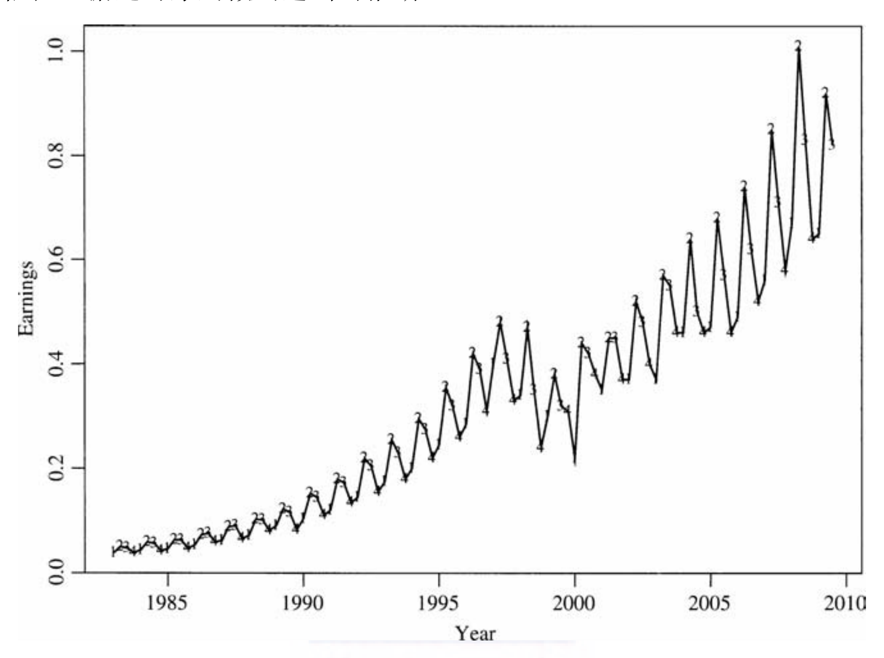


图 2-2　可口可乐公司在 1983 年第 1 季度到 2009 年第 3 季度每股股票的季度盈利

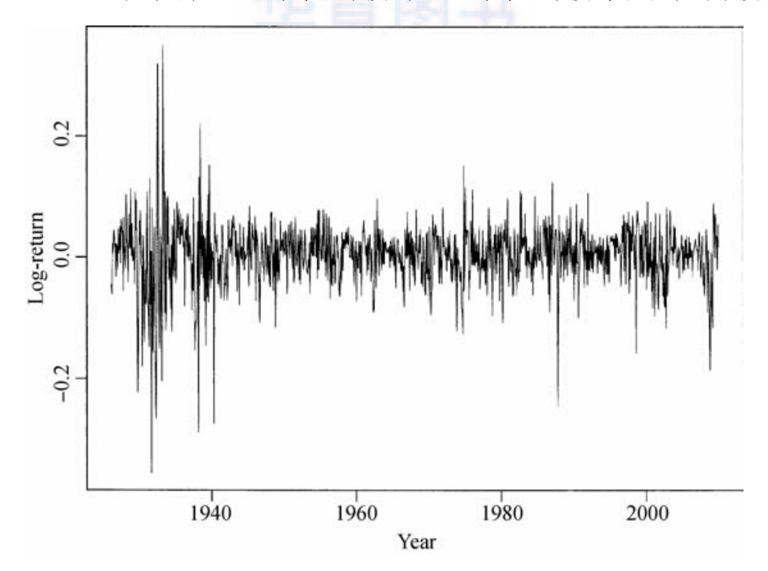


图 2-3　标普 500 指数在 1926 年 1 月到 2009 年 12 月间的月对数收益率

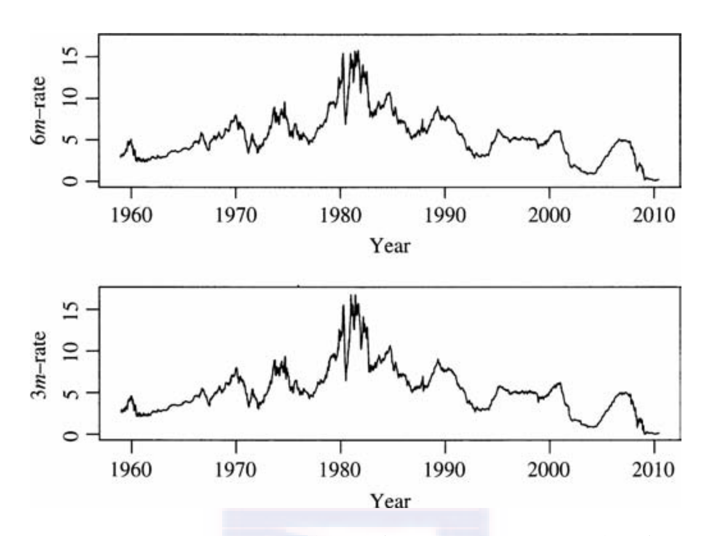


图 2-4 上图是 6 月期国债的利率，下图是 3 月期国债利率

2.1 平稳性

在时间序列分析中，统计推断的基础是弱平稳性的概念. 如图 2-3 中给出的标普 500 指数序列，该序列随时间在 0 值上下变化. 事实上，可以把整个时间区间划分为几个子区间，这些子区间的样本均值都应该接近 0 值. 在统计上，这种现象表明收益率的均值不随时间变化，或者简单地说，期望收益率具有时间不变性. 而且，图 2-3 也说明，在样本数据时间跨度内，除了大萧条时期外，月对数收益率的范围大约在区间[−0.2，0.2]. 在统计上，该特征表明对数收益率的方差不随时间变化. 把这两个时间不变性特征结合在一起，我们称对数收益率序列$\{x_t\}$为**弱平稳的**(weakly stationary). 正式地说，如果一个时间序列 x_t 的一阶矩和二阶矩(即均值和方差)具有时间不变性，则称它为弱平稳的. 弱平稳性是很重要的，因为它为预测提供了基础框架. 对于标普 500 指数的月对数收益率，我们有充分的理由预测未来的月收益率大约在 0 值左右，并且在−0.2～0.2 之间变化.

另一方面，考虑图 2-2 中可口可乐公司每股股票的季度盈利数据，如果把取样时间区间分割为几个子区间，这些子区间上的样本均值就有显著的不同. 所以，该盈利序列不是弱平稳的. 这里没有令人惊讶之处，因为一家优秀公司的季度盈利随时间而提高是符合我们的预期的. 图 2-2 的时序图也表明盈利的波动性随时间而增加. 故此，季度盈利的方差也是随时间变化的. 基于以上原因，季度盈利序列不是弱平稳的. 我们将在后面讨论对这种非平稳序列进行建模的方法和模型.

在讨论弱平稳性时，我们隐含地假定了 x_t 的前两阶矩是有限的. 为了容易表述，我们把 x_t 的均值和方差分别记为$E(x_t)=\mu$ 和$E(x_t-\mu)^2=\gamma_0$. 由于这两个统计量是常数，所以它们不随时间变化，可以分别用两个简单记号 μ 和 γ_0 来表示.

40～42

平稳时间序列 x_t 有其他重要的性质. 对给定的整数 k，称协方差 $\gamma_k=\mathrm{Cov}(x_t，x_{t-k})$为

x_t 的间隔为 k 的自协方差. 应用柯西-施瓦茨(Cauchy-Schwarz)不等式，容易证明 γ_k 存在并且也具有时间不变性. 也就是说，对一个弱平稳的时间序列 x_t，γ_k 仅仅依赖于 k，该值衡量 x_t 和 x_{t-k} 的线性相依性. 或者说，它衡量 x_t 对它的过去取值 x_{t-k} 的动态相依性. 线性时间序列分析着重研究序列 x_t 的动态相依性. 自协方差 γ_k 具有两个重要性质：1) $\gamma_0=\mathrm{Var}(x_t)$；2) $\gamma_{-k}=\gamma_k$. 第二个性质成立，因为 $\gamma_{-k}=\mathrm{Cov}(x_t,x_{t-(-k)})=\mathrm{Cov}(x_{t-(-k)},x_t)=\mathrm{Cov}(x_{t+k},x_t)=\mathrm{Cov}(x_{t_1},x_{t_1-k})=\gamma_k$，其中 $t_1=t+k$.

2.2 相关系数和自相关函数

图 2-5 给出了 IBM 股票的月简单收益率和标普 500 指数的散点图. 这两个收益率看起来是正相关的. 线性相关的程度经常用**皮尔逊**(Pearson)相关系数(或者简称为相关系数)来衡量. 在统计上，两个随机变量 X 和 Y 的相关系数定义为

$$\rho_{x,y}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}=\frac{E[(X-\mu_x)(Y-\mu_y)]}{\sqrt{E(X-\mu_x)^2E(Y-\mu_y)^2}}$$

43

其中 μ_x 和 μ_y 分别表示 X 和 Y 的均值，并且假定方差是有限的. 这个系数度量的是 X 和 Y 线性相依的程度，可以证明 $-1\leqslant\rho_{x,y}\leqslant1$ 且 $\rho_{x,y}=\rho_{y,x}$. 若 $\rho_{x,y}=0$，则这两个随机变量是不相关的. 另外，如果 X 和 Y 都是正态随机变量，则 $\rho_{x,y}=0$ 当且仅当 X 和 Y 是相互独立的.

当我们有样本 $\{(x_t,y_t)\mid t=1,\cdots,T)\}$ 时，相关系数可以由它对应的样本相关系数相合地估计出来：

$$\hat{\rho}_{x,y}=\frac{\sum_{t=1}^{T}(x_t-\overline{x})(y_t-\overline{y})}{\sqrt{\sum_{t=1}^{T}(x_t-\overline{x})^2\sum_{t=1}^{T}(y_t-\overline{y})^2}}$$

其中 $\overline{x}=\sum_{t=1}^{T}x_t/T$ 和 $\overline{y}=\sum_{t=1}^{T}y_t/T$ 分别是 X 和 Y 的样本均值.

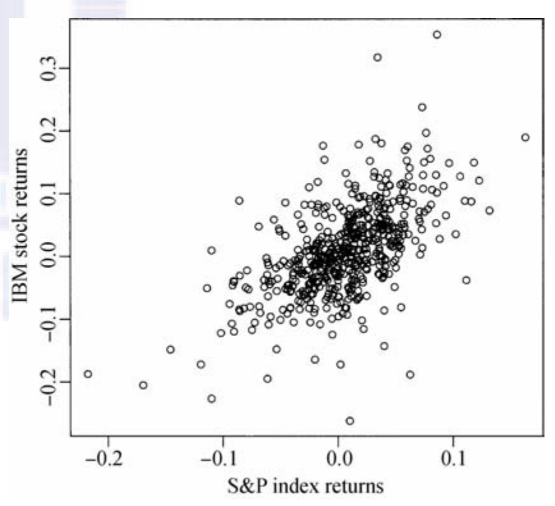


图 2-5 IBM 股票的标普 500 指数的月度简单收益率的散点图. 时间区间是从 1967 年 1 月到 2009 年 12 月

注记：理论上，皮尔逊相关系数取值为 $-1\sim1$ 之间. 然而，对某些随机变量而言，该系数的实际范围比较小. 文献中也给出了其他的衡量相依性的度量指标，其中最流行的两个是斯皮尔曼(Spearman)rho 和肯德尔(Kendall)tau. 因为斯皮尔曼 rho 是基于边际变量秩的相关系数，所以它又称为斯皮尔曼秩相关系数. 肯德尔 tao 则反映了一致数对和非一致数对的差别. 假设 (X_1,Y_1) 和 (X_2,Y_2) 是两个独立同分布的二元连续随机变量. 肯德尔 tau 的定义如下：

$$\tau=P[(X_1-X_2)(Y_1-Y_2)>0]-P[(X_1-X_2)(Y_1-Y_2)<0]$$

为了说明上面的概念，考虑图 2-5 中的散点图. 两个月简单收益的皮尔逊相关系数为

0.5857. 斯皮尔曼 rho 为 0.5861，而肯德尔 tau 为 0.4196. ■

R 代码演示

```
> da=read.table("m-ibmsp6709.txt", header=T)
> head(da)
      date      ibm       sp
1 19670131 0.075370 0.078178
 ....
6 19670630 0.067024 0.017512
> ibm=da$ibm
> sp5=da$sp
> cor(sp5,ibm)
[1] 0.5856544
> cor(sp5,ibm,method='spearman')
[1] 0.5860817
> cor(sp5,ibm,method='kendall')
[1] 0.4196587
```

自相关函数

考虑弱平稳时间序列 x_t. x_t 与 x_{t-k} 的相关系数称为 x_t 的**间隔为 k 的自相关**(lag-k autocorrelation)**系数**，通常记为 ρ_k. 具体地说，其定义为

$$\rho_k = \frac{\mathrm{Cov}(x_t, x_{t-k})}{\sqrt{\mathrm{Var}(x_t)\mathrm{Var}(x_{t-k})}} = \frac{\mathrm{Cov}(x_t, x_{t-k})}{\mathrm{Var}(x_t)} = \frac{\gamma_k}{\gamma_0} \tag{2-1}$$

因为 x_t 为弱平稳的，所以这里有 $\mathrm{Var}(x_{t-k}) = \mathrm{Var}(x_t)$. 根据定义，我们有 $\rho_0 = 1$，$\rho_\ell = \rho_{-\ell}$ 和 $-1 \leqslant \rho_\ell \leqslant 1$. 自相关系数组成的集合$\{\rho_k\}$称为 x_t 的**自相关函数**(Autocorrelation Function，ACF). 一个弱平稳时间序列 x_t 是序列自身前后不相关的，当且仅当对所有 $k > 0$ 都有 $\rho_k = 0$.

对一个给定的样本$\{x_t | t=1, \cdots, T\}$，设$\bar{x}$为样本均值，即 $\bar{x} = \sum_{t=1}^{T} x_t / T$，则 x_t 的间隔为 1 的样本自相关系数为

$$\hat{\rho}_1 = \frac{\sum_{t=2}^{T} (x_t - \bar{x})(x_{t-1} - \bar{x})}{\sum_{t=1}^{T} (x_t - \bar{x})^2}$$

在通常条件下，$\hat{\rho}_1$ 是 ρ_1 的相合估计. 例如，若$\{x_t\}$是独立同分布随机变量序列且 $E(x_t^2) < \infty$，则$\hat{\rho}_1$ 渐近地服从均值为 0、方差为 $1/T$ 的正态分布(见 Brockwell 和 Davis(2009)中的定理 7.2.2). 一般地，x_t 的间隔为 k 的样本自相关系数定义为

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T} (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \quad 0 \leqslant k < T-1 \tag{2-2}$$

若$\{x_t\}$是一个独立同分布随机变量序列，满足 $E(x_t^2) < \infty$，则对任意固定的正整数 k，$\hat{\rho}_k$ 渐近服从均值为 0、方差为 $1/T$ 的正态分布. 更一般地，若 x_t 是一个弱平稳序列，满足 $x_t = \mu + \sum_{i=0}^{q} \psi_i a_{t-i}$，其中 $\psi_0 = 1$，$\{a_j\}$是均值为 0 的独立同分布随机变量序列，则对于 $k > q$，$\hat{\rho}_k$

渐近服从均值为 0、方差为 $\left(1+2\sum_{i=1}^{q}\rho_i^2\right)/T$ 的正态分布．在时间序列文献中这个结果称为 Bartlett 公式(参见 Box 等(1994))．关于样本自相关函数的渐近分布的更多信息，参见 Fuller(1995，第 6 章)和 Brockwell 和 Davis(2009)．

例 2.1 考虑 CRSP 的第 10 分位组合的月简单收益率，时间为 1967 年 1 月到 2009 年 12 月．共计 516 个观测值，即 $T=516$．该投资组合由 NYSE/AMEX/NASDAQ 市值最小的 10%股票构成，并且每年都重新平衡．图 2-6a 给出了收益率序列的时序图，而图 2-6b是序列的样本 ACF．在 ACF 图中的两条水平线代表 2 个标准误差的上下限，即 $\pm 2\sqrt{T}$．ACF 图从 $\hat{\rho}_0=1$ 开始．图 2-6 清楚地表明，在 5%的水平上间隔为 1 的 ACF 是显著不等于 0 的．分析中用到的 R 命令如下：

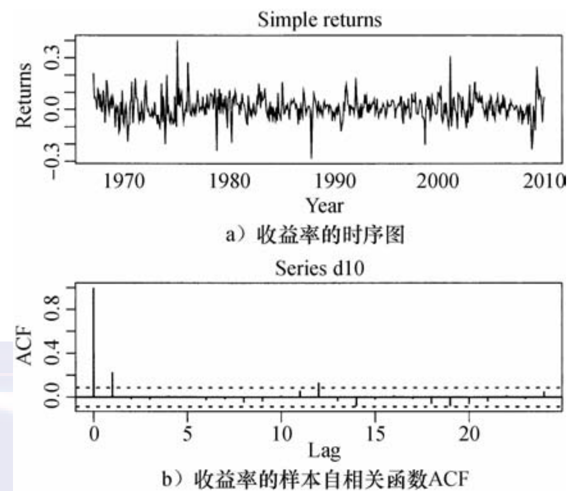


图 2-6 1967 年 1 月到 2009 年 12 月的 CRSP 第 10 分位组合的月简单收益率

44～45

```
> da=read.table("m-dec12910.txt",header=T)
> head(da)
      date      dec1      dec2      dec9     dec10
1 19670131  0.068568  0.080373  0.180843 0.211806
2 19670228  0.008735  0.011044  0.048767 0.064911
> d10=da$dec10   # select the Decile 10 returns
> dec10=ts(d10,frequency=12,start=c(1967,1))
> par(mfcol=c(2,1))
> plot(dec10,xlab='year',ylab='returns')
> title(main='(a): Simple returns')
> acf(d10,lag=24) # command to obtain sample ACF of the data
```

■

检验单个 ACF

对一个给定的正整数 k，可用前面的结果来检验 $H_0：\rho_k=0$ 对 $H_a：\rho_k\neq 0$．检验统计量为：

$$t\text{ 比}=\frac{\hat{\rho}_k}{\sqrt{\left(1+2\sum_{i=1}^{k-1}\hat{\rho}_i^2\right)/T}}$$

如果$\{x_t\}$是平稳的高斯序列并且满足当 $j>k$ 时 $\rho_j=0$，则该 t 比(t-ratio)渐近地服从标准正态分布．所以，决策规则是：当$|t\text{ 比}|>Z_{\alpha/2}$时拒绝 H_0，其中 $Z_{\alpha/2}$是标准正态分布的 $100(1-\alpha/2)$百分位分位点．另外，也可以用上述统计量 t 比的 p 值来得出结论．如果 p 值小于第 I 类错误值，例如 0.05，那么就拒绝原假设．如果 p 值大于或者等于第 I 类错误值，那么就不能拒绝原假设 H_0．

46

为了简单起见，对于任意 $k\neq 0$，许多软件用 $1/T$ 作为$\hat{\rho}_k$ 的渐进方差．那么 t 比就变为

$\sqrt{T}\hat{\rho}_k$. 该简单近似实质上是假定所分析的时间序列是一个独立同分布随机变量序列.

例 2.1(续) 在美国，小市值股票倾向在 1 月有一个正收益率，这可能由于多种原因，例如税收考虑，或者年终的组合调整. 这称为小市值股票的 1 月效应(January effect). 一种验证小市值股票 1 月效应存在性的方法是应用例 2.1 中 CRSP 组合的最低 10 分位的月简单收益率来进行假设检验：

$$\text{原假设 } H_0: \rho_{12}=0 \text{ 和备择假设 } H_a: \rho_{12}\neq 0$$

从数据中，我们计算得到$\hat{\rho}_{12}=0.13$. 应用 $1/\sqrt{T}$作为$\hat{\rho}_{12}$的标准误差的渐进估计，我们得到 t 比 $t=\sqrt{T}\hat{\rho}_{12}=2.96$，它大于 5%的临界值 1.96. 所以，我们拒绝原假设，即在 5%显著性水平下 $\rho_{12}=0$. 换句话说，数据证实了小市值股票收益率 1 月效应的存在性.

R 代码演示

```
> f1=acf(d10,lag=24)
> f1$acf
 [1,]   1.000000000    # lag-0
 [2,]   0.227386585    # lag-1
   .    .......
[13,]   0.130411045    # lag-12
[14,] -0.036881195    # lag-13
> tt=f1$acf[13]*sqrt(516)
> tt
[1] 2.962369
```

■

对有限样本，$\hat{\rho}_k$ 是 ρ_k 的有偏估计. 偏差的阶是 $1/T$，这在样本容量 T 较小的时候是不容忽视的. 但在大多数金融应用中，T 相对较大，故这个偏差并不严重.

混成检验

由式(2-2)所定义的统计量$\hat{\rho}_1$，$\hat{\rho}_2$，…称为 x_t 的样本**自相关函数**(ACF). 这个函数在线性时间序列分析中起着重要作用. 事实上，一个线性时间序列模型可以完全由其 ACF 来刻画，并且线性时间序列的建模用样本 ACF 来决定数据的线性动态关系. 在许多金融应用中，我们经常需要联合检验 x_t 的多个自相关系数是否同时为 0. Box 和 Pierce(1970)提出了混成统计量(Portmanteau statistic)

$$Q_*(m) = T\sum_{\ell=1}^{m}\hat{\rho}_\ell^2$$

47

来检验原假设 $H_0: \rho_1=\cdots=\rho_m=0$ 对备择假设 $H_a: \rho_i\neq 0(i\in\{1,\cdots,m\})$. 在$\{x_t\}$是满足一定矩条件的 iid 随机序列的假定下，$Q_*(m)$渐近地是自由度为 m 的 χ^2 随机变量.

为了提高有限样本中检验的功效，Ljung 和 Box(1978)把统计量 $Q_*(m)$修改成

$$Q(m) = T(T+2)\sum_{\ell=1}^{m}\frac{\hat{\rho}_\ell^2}{T-\ell} \tag{2-3}$$

决策规则是：当 $Q(m)>\chi_\alpha^2$ 时拒绝 H_0，其中 χ_α^2 是自由度为 m 的 χ^2 分布的 $100(1-\alpha)$百分位分位点. 大部分软件包都会给出 $Q(m)$的 p 值. 决策规则是当 p 值小于第 I 类错误或者显著性水平 α 时，拒绝 H_0.

例 2.2 考虑 IBM 股票从 1967 年 1 月到 2009 年 12 月的月简单收益率和对数收益率. 样本容量是 516. 图 2-7 给出了 IBM 股票的简单月收益率和对数收益率的样本自相关函数. 这两个样本自相关函数非常接近，它们都在 2 倍的标准误差范围内，这表明 IBM 股票月收益率序列即使存在某种相关性，该自相关性也是很小的. 为了验证该收益率序列没有序列自相关性，我们检验 $H_0: \rho_1=\rho_2=\cdots=\rho_m=0$ 对 $H_a: \rho_i\neq 0(i\in\{1, ...m\})$，其中 $m=12$ 和 24. 对于简单收益率，我们有 $Q(12)=7.57$，其 p 值为 0.82；$Q(24)=25.49$，其对应的 p 值为 0.38. 对于对数收益率，$Q(12)=7.40$，$Q(24)=25.49$，它们对应的 p 值分别为 0.83 和 0.38. 因此，Ljung-Box 统计量不能拒绝 IBM 股票收益率没有显著的前后相关性这一原假设.

R 代码演示

```
> da=read.table("m-ibmsp6709.txt",header=T)
> ibm=da$ibm
> lnibm=log(ibm+1) # Transfer to log returns
> Box.test(ibm,lag=12,type='Ljung')
        Box-Ljung test
data:  ibm
X-squared = 7.5666, df = 12, p-value = 0.818

> Box.test(lnibm,lag=12,type='Ljung')
        Box-Ljung test
data:  lnibm
X-squared = 7.4042, df = 12, p-value = 0.8298
```

■

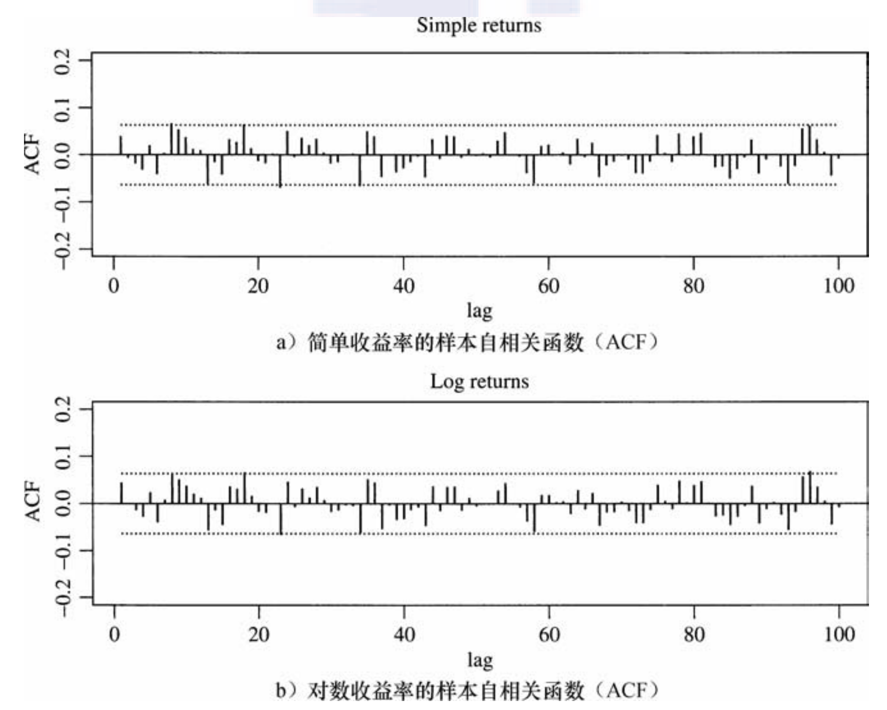


图 2-7 从 1967 年 1 月到 2009 年 12 月的 IBM 股票的月简单收益率和对数收益率的样本自相关函数

实际上，由于 m 的选择会影响 $Q(m)$ 的表现，所以经常选用多个 m 值. 模拟研究表明 $m \approx \ln(T)$ 会有较好的功效. 在分析季节性时间序列时，由于时间隔为周期的倍数的自相关系数更加重要，所以这个一般性的规则需要加以修正.

例 2.1(续) 对例 2.1 中 CRSP 组合的最低 10 分位的资产组合的月简单收益率，Ljung-Box 统计量表明 $Q(12)=41.06$，对应的 p 值为 4.79×10^{-5}；$Q(24)=56.25$，对应的 p 值为 2.12×10^{-4}. 这些 p 值都很小，所以在 5%显著性水平下，序列不存在自相关这一原假设被拒绝. 所以，最低 10 分位资产组合的月简单收益率存在序列自相关性. ■

设 x_t 为金融资产的收益率序列. 在金融文献中，资本资产定价模型(Capital Asset Pricing Model，CAPM)理论的一种形式就假定序列 $\{x_t\}$ 是不可预测的，且没有自相关性. 自相关系数是否为 0 的检验用来作为有效市场假定是否成立的判定工具. 然而，股价的决定方式和指数收益率的计算方式可能会导致在观测到的收益率序列中有自相关性. 这在分析高频金融数据时尤其常见. 在第 6 章中我们将讨论这方面的问题，例如买-卖价格的跳跃与非同步交易.

48 ~ 49

2.3 白噪声和线性时间序列

白噪声

如果时间序列 x_t 是一个具有有限均值和有限方差的独立同分布随机变量序列，则 $\{x_t\}$ 称为一个**白噪声序列**(white noise). 特别地，若 x_t 还服从均值为 0、方差为 σ^2 的正态分布，则称这个序列为**高斯白噪声**(Gaussian white noise). 对于白噪声序列，所有自相关函数为零. 在实际应用中，如果所有样本自相关函数接近于零，则认为该序列是白噪声序列. 基于图 2-7 和图 2-6b，IBM 股票的月收益率是接近白噪声的，而最低 10 分位资产组合的收益率不是白噪声.

下面我们讨论一些简单的统计模型，它们对时间序列的动态结构建模非常有用，而且这里所述的概念在后面资产收益率的波动率建模中也是很有用的.

线性时间序列

时间序列 x_t 称为线性序列，如果它能写成

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i} \tag{2-4}$$

其中 μ 是 x_t 的均值，$\psi_0=1$，$\{a_t\}$ 是零均值独立同分布并且有合理定义的分布的随机变量序列(也即 $\{a_t\}$ 是白噪声序列). 我们在以后可以看出，a_t 表示时间序列在 t 时刻出现的新信息，因此常将 a_t 称为时刻 t 的**新息**(innovation)或**扰动**(shock). 本书中我们主要关心 a_t 为连续型随机变量的情形. 并不是所有金融时间序列都是线性的，但是实际应用中线性模型通常可以提供精确的近似.

在式(2-4)定义的线性时间序列中，系数 ψ_i 决定了 x_t 的动态结构，在时间序列文献中这些系数称为 x_t 的 ψ 权重(ψ-weight). 若 x_t 是弱平稳的，我们利用 $\{a_t\}$ 的性质可以很容易

地得到 x_t 的均值和方差

$$E(x_t)=\mu,\quad \mathrm{Var}(x_t)=\sigma_a^2\sum_{i=0}^{\infty}\psi_i^2 \tag{2-5}$$

其中 σ_a^2 是 a_t 的方差. 因为 $\mathrm{Var}(x_t)<\infty$，所以 $\{\psi_i^2\}$ 必须是收敛序列，即当 $i\to\infty$ 时 $\psi_i^2\to 0$. 因此，对于一个平稳序列，随着 i 的增大遥远的扰动 a_{t-i} 对 x_t 的影响会逐渐消失.

x_t 的间隔为 ℓ 的自协方差为

50

$$\begin{aligned}\gamma_\ell=\mathrm{Cov}(x_t,x_{t-\ell})&=E\left[\left(\sum_{i=0}^{\infty}\psi_i a_{t-i}\right)\left(\sum_{j=0}^{\infty}\psi_j a_{t-\ell-j}\right)\right]\\&=E\left(\sum_{i,j=0}^{\infty}\psi_i\psi_j a_{t-i}a_{t-\ell-j}\right)\\&=\sum_{j=0}^{\infty}\psi_{j+\ell}\psi_j E(a_{t-\ell-j}^2)\\&=\sigma_a^2\sum_{j=0}^{\infty}\psi_j\psi_{j+\ell}\end{aligned} \tag{2-6}$$

因此，ψ 权重与 x_t 的自相关系数有如下关系：

$$\rho_\ell=\frac{\gamma_\ell}{\gamma_0}=\frac{\sum_{i=0}^{\infty}\psi_i\psi_{i+\ell}}{1+\sum_{i=1}^{\infty}\psi_i^2}\quad \ell\geqslant 0 \tag{2-7}$$

其中 $\psi_0=1$. 线性时间序列模型就是用来描述 x_t 的 ψ 权重的计量模型和统计模型. 对弱平稳时间序列而言，当 $i\to\infty$ 时 $\psi_i\to 0$，从而随着 ℓ 的增加 ρ_ℓ 收敛到 0. 对于资产收益率而言，这意味着，如所期望的那样，当 ℓ 较大时，当前收益率 x_t 对遥远的过去收益率 $x_{t-\ell}$ 的线性相依会消失.

2.4 简单自回归模型

当 x_t 具有统计显著的间隔为 1 的自相关系数时，滞后值 x_{t-1} 可能会在预测 x_t 时有用. 下面的简单模型可以利用这样的预测功用：

$$x_t=\phi_0+\phi_1 x_{t-1}+a_t \tag{2-8}$$

其中 $\{a_t\}$ 是均值为 0、方差为 σ_a^2 的白噪声序列. 这个模型与众所周知的简单线性回归模型有相同的形式，这里 x_t 是因变量，x_{t-1} 是解释变量. 在时间序列文献中，模型(式(2-8))称为一阶**自回归(AR)模型**，或简称 AR(1)模型. 该模型也广泛地应用在随机波动率的建模中，只不过 x_t 由它的对数波动率所代替(见第 4 章).

式(2-8)中的 AR(1)模型有若干类似于简单线性回归模型的性质. 但是，这两个模型之间存在一些显著的差异，这一点将在以后讨论. 这里我们注意这样一个事实就够了：在已知过去收益率 x_{t-1} 的条件下，由 AR(1)模型可推得：

$$E(x_t|x_{t-1})=\phi_0+\phi_1 x_{t-1},\quad \mathrm{Var}(x_t|x_{t-1})=\mathrm{Var}(a_t)=\sigma_a^2$$

也就是说，给定过去收益率 x_{t-1}，本期收益率将以 $\phi_0+\phi_1 x_{t-1}$ 为中心取值，标准差是 σ_a. 给定 x_{t-1} 条件下，收益率 x_t 与 x_{t-i} 无关，其中 $i>1$，这是 AR(1)的马尔科夫性. 显然，某

51

些情况下 x_{t-1} 不能单独决定 x_t 的条件期望，所以需要更灵活一些的模型．AR(1)模型的直接推广是 AR(p)模型：

$$x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + a_t \tag{2-9}$$

其中 p 是非负整数，$\{a_t\}$的定义与式(2-8)中一样．这个模型说的是给定过去数据，过去的 p 个值 x_{t-i}($i=1$，…，p)联合决定 x_t 的条件期望．AR(p)模型与以 p 个滞后值作为解释变量的多元线性回归有相同形式．

2.4.1　AR 模型的性质

为了有效地利用 AR 模型，有必要对它的基本性质进行研究．我们详细地讨论 AR(1)和 AR(2)模型的性质，对一般的 AR(p)模型只给出结果．

AR(1)模型

我们首先讨论式(2-8)定义的 AR(1)模型弱平稳性的充分必要条件．假定序列是弱平稳的，则 $E(x_t)=\mu$，$\mathrm{Var}(x_t)=\gamma_0$，$\mathrm{Cov}(x_t, x_{t-j})=\gamma_j$，其中 μ 和 γ_0 是常数，γ_j 是 j 的函数而与 t 无关．我们容易得到序列的均值、方差和自相关系数．对式(2-8)两边取期望，因为 $E(a_t)=0$，所以

$$E(x_t) = \phi_0 + \phi_1 E(x_{t-1})$$

在平稳性的条件下，$E(x_t)=E(x_{t-1})=\mu$，从而

$$\mu = \phi_0 + \phi_1 \mu \quad 或 \quad E(x_t) = \mu = \frac{\phi_0}{1-\phi_1}$$

这个结果对 x_t 有两个含义：第一，若 $\phi_1 \neq 1$，则 x_t 的均值存在；第二，x_t 的均值为 0 当且仅当 $\phi_0=0$．因此，对平稳 AR(1)过程，常数项 ϕ_0 与 x_t 的均值有关，$\phi_0=0$ 意味着 $E(x_t)=0$．

我们利用 $\phi_0=(1-\phi_1)\mu$ 可以把 AR(1)模型写成如下形式

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + a_t \tag{2-10}$$

重复代入，由上述方程可推得

$$\begin{aligned} x_t - \mu &= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots \\ &= \sum_{i=0}^{\infty} \phi_1^i a_{t-i} \end{aligned} \tag{2-11}$$

如果设 $\psi_i=\phi_1^i$，上式就把 AR(1)模型表达为式(2-4)的形式．因此，$x_t-\mu$ 是 a_{t-i} 的线性函数，其中 $i \geqslant 0$．利用这个性质和$\{a_t\}$的独立性，我们有 $E[(x_t-\mu)a_{t+1}]=0$．根据平稳性的假定，我们有 $\mathrm{Cov}(x_{t-1}, a_t)=E[(x_{t-1}-\mu)a_t]=0$．因为 x_{t-1} 发生在 t 时刻之前，而 a_t 作为 t 时刻的扰动，它不依赖于任何过去的信息，所以上面等式的后一个结果可以从这些事实得出来．对式(2-10)两边平方，然后取期望，得到

$$\mathrm{Var}(x_t) = \phi_1^2 \mathrm{Var}(x_{t-1}) + \sigma_a^2$$

其中 σ_a^2 是 a_t 的方差，这里我们用到"x_{t-1} 与 a_t 的协方差为 0"这样一个事实．而在平稳性的假定下，$\mathrm{Var}(x_t)=\mathrm{Var}(x_{t-1})$，因此

$$\mathrm{Var}(x_t) = \frac{\sigma_a^2}{1-\phi_1^2}$$

在 $\phi_1^2<1$ 时成立．$\phi_1^2<1$ 的要求源于随机变量方差的非负性和 x_t 的弱平稳性．因此，由

AR(1)模型的弱平稳性可推出$-1<\phi_1<1$，即$|\phi_1|<1$. 反之，若$|\phi_1|<1$，由式(2-11)和序列$\{a_t\}$的独立性，可以证明x_t的均值和方差是有限的并且不随时间变化，参见式(2-5). 另外，由式(2-6)，x_t的自协方差也是有限的. 所以，AR(1)模型是弱平稳的. 综上所述，式(2-8)定义的AR(1)模型是弱平稳的充分必要条件是$|\phi_1|<1$.

应用$\phi_0=(1-\phi_1)\mu$，可以把AR(1)模型重新写为：

$$x_t = (1-\phi_1)\mu + \phi_1 x_{t-1} + a_t$$

在金融文献中，这个模型常用ϕ_1来刻画AR(1)时间序列的动态相依性的持续性.

AR(1)模型的自相关函数

在式(2-10)两端乘以a_t，再取期望，利用a_t与x_{t-1}的独立性，我们有

$$E[a_t(x_t-\mu)] = \phi_1 E[a_t(x_{t-1}-\mu)] + E(a_t^2) = E(a_t^2) = \sigma_a^2$$

其中σ_a^2是a_t的方差. 对式(2-10)两端同乘以$(x_{t-\ell}-\mu)$，取期望，再利用上述结果，我们有

$$\gamma_\ell = \begin{cases} \phi_1\gamma_1 + \sigma_a^2 & \ell = 0 \\ \phi_1\gamma_{\ell-1} & \ell > 0 \end{cases}$$

这里我们用了$\gamma_\ell=\gamma_{-\ell}$这个性质. 因此，对式(2-8)定义的弱平稳AR(1)模型，我们有

$$\mathrm{Var}(x_t) = \gamma_0 = \frac{\sigma^2}{1-\phi_1^2} \text{ 和 } \gamma_\ell = \phi_1\gamma_{\ell-1} \quad \ell > 0$$

53

由后一个方程，x_t的自相关函数(ACF)满足

$$\rho_\ell = \phi_1\rho_{\ell-1} \quad \ell > 0$$

因为$\rho_0=1$，所以有$\rho_\ell=\phi_1^\ell$. 这个性质表明弱平稳AR(1)序列的自相关函数从$\rho_0=1$开始以比率为ϕ_1的指数速度衰减. 对正的ϕ_1，AR(1)模型的自相关函数(ACF)图像呈现漂亮的指数衰减. 对负的ϕ_1，AR(1)模型的ACF由上下两个都以ϕ_1^2比率衰减的图像组成. 图2-8显示，当$\phi_1=0.8$和$\phi_1=-0.8$时的AR(1)模型的ACF图.

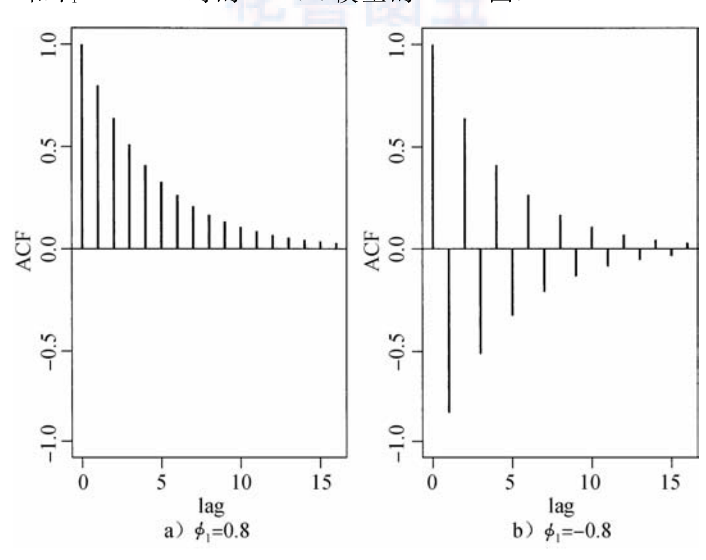


图2-8　AR(1)模型的自相关函数

AR(2)模型

AR(2)模型为

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + a_t \tag{2-12}$$

利用与 AR(1)模型相同的方法，我们得到：只要 $\phi_1+\phi_2\neq 1$，就有

$$E(x_t) = \mu = \frac{\phi_0}{1-\phi_1-\phi_2}$$

利用 $\phi_0=(1-\phi_1-\phi_2)\mu$，我们可把 AR(2)改写成

54

$$x_t - \mu = \phi_1(x_{t-1}-\mu) + \phi_2(x_{t-2}-\mu) + a_t$$

在上式两端同乘以 $x_{t-\ell}-\mu$，我们有

$$(x_{t-\ell}-\mu)(x_t-\mu) = \phi_1(x_{t-\ell}-\mu)(x_{t-1}-\mu) + \phi_2(x_{t-\ell}-\mu)(x_{t-2}-\mu) + (x_{t-\ell}-\mu)a_t$$

对上式两边取期望，并利用当 $\ell>0$ 时 $E[(x_{t-\ell}-\mu)a_t]=0$ 这个性质，我们得到

$$\gamma_\ell = \phi_1\gamma_{\ell-1} + \phi_2\gamma_{\ell-2} \quad \ell > 0$$

这个结果称为平稳 AR(2)模型的*矩方程*(moment equation)，在上式两端同除以 γ_0，得到 x_t 的 ACF 的性质：

$$\rho_\ell = \phi_1\rho_{\ell-1} + \phi_2\rho_{\ell-2} \quad \ell > 0 \tag{2-13}$$

特别地，间隔为 1 的 ACF 满足

$$\rho_1 = \phi_1\rho_0 + \phi_2\rho_{-1} = \phi_1 + \phi_2\rho_1$$

因此，对平稳的 AR(2)序列 x_t，我们有 $\rho_0=1$，

$$\rho_1 = \frac{\phi_1}{1-\phi_2}$$

$$\rho_\ell = \phi_1\rho_{\ell-1} + \phi_2\rho_{\ell-2} \quad \ell \geqslant 2$$

式(2-13)的结果说的是：平稳 AR(2)序列的 ACF 满足二阶差分方程

$$(1-\phi_1 B - \phi_2 B^2)\rho_\ell = 0$$

其中 B 是**延迟算子**(backshift)(或滞后算子)，即 $B\rho_\ell=\rho_{\ell-1}$. 这个差分方程决定了平稳 AR(2)序列的 ACF 性质，也决定了 x_t 的预测行为. 在时间序列的文献中，有时也用 L 而不是 B 来表示延迟算子. 可以用 L 表示滞后(lag)算子，例如 $Lx_t=x_{t-1}$，$L\psi_k=\psi_{k-1}$.

与上面的差分方程相对应的是二次多项式方程

$$1-\phi_1 z - \phi_2 z^2 = 0 \tag{2-14}$$

这个方程的解是

$$z = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$$

55

在时间序列文献中，称这两个解的倒数为 AR(2)模型的**特征根**(characteristic root). 用 ω_1 和 ω_2 表示这两个特征根. 如果两个 ω_i 都是实值的，则模型的二阶差分方程能分解为$(1-\omega_1 B)(1-\omega_2 B)$，这时 AR(2)模型可以看成是一个 AR(1)模型在另一个 AR(1)模型上的运算. 因此，这时 x_t 的 ACF 是两个指数衰减的混合. 然而，如果 $\phi_1^2+4\phi_2<0$，则 ω_1 和 ω_2 都是复数(称为**复共轭对**(complex conjugate pair))，这时 x_t 的 ACF 图形将呈现出减幅的正弦和余弦波图像. 在商业和经济的应用中，复特征根是很重要的，它们可以刻画商业周

期的行为. 对经济时间序列模型来说，复特征根是常见的. 对于式(2-12)定义的带一对共轭复特征根的 AR(2)模型，随机周期的**平均**(average)长度为

$$k=\frac{2\pi}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]}$$

其中反余弦函数的值是以弧度来表示的. 如果我们将复数解写成 $a\pm b\mathrm{i}$ 的形式，其中 $\mathrm{i}=\sqrt{-1}$，则我们有 $\phi_1=2a$，$\phi_2=-(a^2+b^2)$，并且

$$k=\frac{2\pi}{\cos^{-1}(a/\sqrt{a^2+b^2})}$$

其中 $\sqrt{a^2+b^2}$ 是 $a\pm b\mathrm{i}$ 的模. 参见例 2.3 的详细说明.

图 2-9 显示的是 4 个平稳 AR(2)模型的 ACF. 图 2-9b 是 AR(2)模型 $(1-0.6B+0.4B^2)x_t=a_t$ 的 ACF. 因为 $\phi_1^2+4\phi_2=0.36+4\times(-0.4)=-1.24<0$，这个 AR(2)模型包含两个复特征根，所以它的 ACF 呈现出减幅的正弦和余弦波状. 其他 3 个 AR(2)模型都是有实特征根的，它们的 ACF 呈指数衰减.

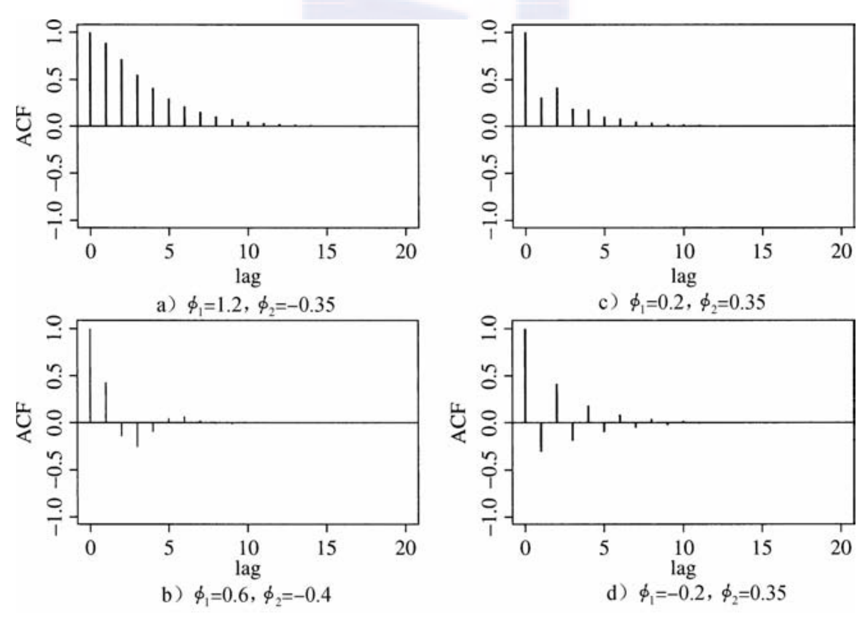


图 2-9　AR(2)模型的自相关函数

例 2.3　作为说明，考虑美国的实际国民生产总值(Gross National Product，GNP)季节调整后的季度增长率，时间是从 1947 年第二季度到 2010 年第一季度，总计 252 个观测值. 图 2-10 给出了该 GNP 序列的对数序列(单位：10 亿美元)和增长率的时序图. 其中图 2-10b 的增长率时序图中增加了一条对应 0 值的水平线. 图中清晰地表明大部分的增长率为正值，最大的 GNP 下跌出现在 2008 年衰退期. 基于下一节中要讲述的建模过程，这里我们对该序列应用一个 AR(3)模型. 拟合的模型是：

$$(1-0.438B-0.206B^2+0.156B^3)(x_t-0.016)=a_t,\quad \hat{\sigma}_a=9.55\times10^{-5}\tag{2-15}$$

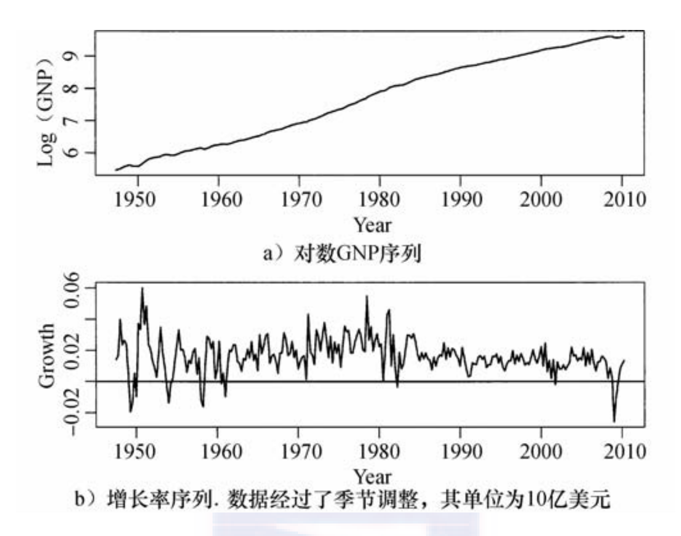


图 2-10　从 1947 年第一季度到 2010 年第一季度的美国 GNP 序列时序图

56～57

模型的标准误差分别为 0.062、0.067、0.063 和 0.001．详细内容可以参见所附的 R 输出结果．模型(式(2-15))给出了一个 3 阶多项式方程，即

$$1-0.438z-0.206z^2+0.156z^3=0$$

该方程有 3 个解，即 1.616+0.864i、1.616−0.864i 和−1.909．其中的实根对应于一个因子$[1-(1/-1.909)z]=(1+0.524z)$，该因子表明了该 GNP 增长率的指数衰减特性．对于复共轭对 1.616±0.864i，我们得到其模为$\sqrt{1.616^2+0.864^2}=1.833$，以及

$$k=\frac{2\pi}{\cos^{-1}(1.616/1.833)}\approx 12.80$$

因此，这个拟合的 AR(3)模型确证了美国经济中存在随机周期．该周期的平均长度大约为 12.8 个季度，这大约为 3 年．该结果是合理的，因为美国经济会经历膨胀期和紧缩期，一般相信膨胀期的长度为 3 年．若我们用一个非线性模型把美国经济分解成“膨胀期”和“紧缩期”的话，数据将表明紧缩期平均持续大约 3 个季度，而膨胀期的平均持续期为 3 年．例如，可以参见 Tsay (2010，第 4 章)．平均持续期 12.8 个季度是这两个平均持续期的折中．这里得到的周期性在国民经济增长率的研究中是常见的，例如类似上面所述的现象在经济合作与发展组织(Organization for Economic Cooperation and Development，OECD，简称为经合组织)国家也是存在的.

```
> da=read.table("q-gnp4710.txt",header=T)
> head(da)
  Year Mon Day VALUE
1 1947   1   1 238.1
 ...
6 1948   4   1 268.7
> G=da$VALUE
> LG=log(G)
> gnp=diff(LG)
> dim(da)
[1] 253   4
```

```
> tdx=c(1:253)/4+1947 # create the time index
> par(mfcol=c(2,1))
> plot(tdx,LG,xlab='year',ylab='GNP',type='l')
> plot(tdx[2:253],gnp,type='l',xlab='year',ylab='growth')
> acf(gnp,lag=12)
> pacf(gnp,lag=12) # compute PACF
> m1=arima(gnp,order=c(3,0,0))
> m1
Call:
arima(x = gnp, order = c(3, 0, 0))

Coefficients:
         ar1     ar2      ar3  intercept
      0.4386  0.2063  -0.1559     0.0163
s.e.  0.0620  0.0666   0.0626     0.0012
sigma^2 estimated as 9.549e-05:log likelihood=808.6,aic=-1607.1
> tsdiag(m1,gof=12)  # model checking discussed later
> p1=c(1,-m1$coef[1:3]) # set-up the polynomial
> r1=polyroot(p1) # solve the polynomial equation
> r1
[1]  1.616116+0.864212i -1.909216-0.000000i  1.616116-0.864212i
> Mod(r1)
[1] 1.832674 1.909216 1.832674  # compute absolute values
> k=2*pi/acos(1.616116/1.832674) # compute length of the period
> k
[1] 12.79523
```

■

平稳性

AR(2)时间序列的平稳性条件是它的两个特征根的绝对值都小于1，或者说，它的两个特征根的模小于1．因此，特征方程两个解的模要大于1．在这种条件下，式(2-13)中的递归式保证模型的自相关函数随间隔ℓ的增加而趋于0，这个趋于0的性质是一个时间序列平稳的必要条件．事实上，这个条件也适用于AR(1)模型，这时特征方程是$1-\phi_1 z=0$．它的特征根为$\omega=1/z=\phi_1$，要使x_t是平稳的，必须有ϕ_1的模小于1．而前面已证明过，对平稳AR(1)模型有$\rho_\ell=\phi_1^\ell$．上面的条件意味着，当$\ell\to\infty$时有$\rho_\ell\to 0$．

AR(p)模型

AR(1)和AR(2)模型的结果可以方便地推广到由式(2-9)定义的AR(p)模型．对平稳的AR(p)序列，其均值为

$$E(x_t)=\frac{\phi_0}{1-\phi_1-\cdots-\phi_p}$$

假定分母不等于0．该模型对应的特征方程为

$$1-\phi_1 z-\phi_2 z^2-\cdots-\phi_p z^p=0$$

如果这个方程所有解的模都大于1，则序列x_t是平稳的．同样，该方程解的倒数为该模型的**特征根**(characteristic root)．因此，平稳性要求所有特征根的模都小于1．对平稳AR(p)序列而言，其自相关函数满足下面的差分方程

$$(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)\rho_\ell=0\quad \ell>0$$

58~59

平稳 AR(p)模型的自相关函数(ACF)图像呈现出减幅的正弦、余弦和指数衰减的混合状，具体形状取决于其特征根的性质.

2.4.2 实践中 AR 模型的识别

在实际应用中，AR 时间序列的阶数 p 是未知的，必须根据实际数据来决定. 这个问题叫做 AR 模型的**定阶**(order determination)，这个问题在时间序列文献中已被广泛研究. 一般有两个确定 p 值的方法：第一种方法是利用偏自相关函数(Partial AutoCorrelation Function，PACF)，第二种方法是用某个信息准则函数.

偏自相关函数(PACF)

平稳时间序列的 PACF 是 ACF 的函数，它在给 AR 模型定阶时是一个有用的工具. 一个简单而有效的引进 PACF 的方式是依次考虑如下一系列的 AR 模型：

$$
\begin{aligned}
x_t &= \phi_{0,1} + \phi_{1,1} x_{t-1} + e_{1t} \\
x_t &= \phi_{0,2} + \phi_{1,2} x_{t-1} + \phi_{2,2} x_{t-2} + e_{2t} \\
x_t &= \phi_{0,3} + \phi_{1,3} x_{t-1} + \phi_{2,3} x_{t-2} + \phi_{3,3} x_{t-3} + e_{3t} \\
x_t &= \phi_{0,4} + \phi_{1,4} x_{t-1} + \phi_{2,4} x_{t-2} + \phi_{3,4} x_{t-3} + \phi_{4,4} x_{t-4} + e_{4t} \\
&\vdots
\end{aligned}
$$

其中，$\phi_{0,j}$是常数项，$\phi_{i,j}$是x_{t-i}的系数，$\{e_{jt}\}$是 AR(j)模型的误差项. 这些模型都是多元线性回归的形式，可用最小二乘法来估计. 事实上，由于它们是按阶的高低排列的，所以我们可以应用多元线性回归分析中偏 F 检验的思想. 第一个式子中的估计$\hat{\phi}_{1,1}$称为 x_t 的间隔为 1 的样本偏自相关函数；第二个式子中的估计$\hat{\phi}_{2,2}$称为 x_t 的间隔为 2 的样本偏自相关函数；第三个式子中的估计$\hat{\phi}_{3,3}$称为 x_t 的间隔为 3 的样本偏自相关函数；以此类推.

从定义可以看出，间隔为 2 的样本偏自相关函数$\hat{\phi}_{2,2}$所表示的是：在 AR(1)模型 $x_t=\phi_0+\phi_1 x_{t-1}+e_{1t}$基础上添加的 x_{t-2}对 x_t 的贡献；$\hat{\phi}_{3,3}$表示的是在 AR(2)模型上添加的 x_{t-3}对 x_t 的贡献等. 因此，对一个 AR(p)模型，间隔为 p 的样本偏自相关函数 PACF 不应为零，而对所有$j>p$，$\phi_{j,j}$应接近于零. 我们利用这一性质来决定阶数 p. 对于平稳高斯 AR(p)模型，可以证明其样本偏自相关函数(PACF)有如下性质：

- 当样本容量 T 趋于无穷大时，$\hat{\phi}_{p,p}$收敛于ϕ_p.
- 对 $\ell>p$，$\hat{\phi}_{\ell,\ell}$收敛于零.
- 对 $\ell>p$，$\hat{\phi}_{\ell,\ell}$的渐近方差为 $1/T$.

60

这些结果表明 AR(p)序列的样本偏自相关函数是 p 步截尾的.

作为例子，考虑从 1926 年 1 月到 2008 年 12 月 CRSP 价值加权指数的月简单收益率. 表 2-1 给出了这个序列的样本自相关函数在前 12 个间隔上的值. $T=996$，该样本偏自相关函数的渐近标准差大约为 0.032. 从而，在 5%的显著性水平下，对该数据我们识别出一个 AR(3)模型或 AR(9)模型(也就是，$p=3$ 或 9). 如果应用 1%的显著性水平，我们则识别一个 AR(3)模型.

表 2-1 从 1926 年 1 月到 2008 年 12 月 CRSP 价值加权指数的月简单收益率的样本偏自相关函数和信息准则 AIC 及 BIC

p	1	2	3	4	5	6
PACF	0.115	−0.030	−0.102	0.033	0.062	−0.050
AIC	−5.838	−5.837	−5.846	−5.845	−5.847	−5.847
BIC	−5.833	−5.827	−5.831	−5.825	−5.822	−5.818

p	7	8	9	10	11	12
PACF	0.031	0.052	0.063	0.005	−0.005	0.011
AIC	−5.846	−5.847	−5.849	−5.847	−5.845	−5.843
BIC	−5.812	−5.807	−5.805	−5.798	−5.791	−5.784

作为另外一个例子，图 2-11 给出了例 2.3 中 GNP 增长率的 PACF．图 2-11 中的两条虚线表示的是大约为两个标准误差的上下限$\pm 2/\sqrt{252}$．该 PACF 图表明 AR(3)模型适合该数据，因为在 5%水平上，样本 PACF 在前 3 个间隔点上看起来是显著的．间隔为 9 的 PACF 也略微显著．而为了简单，我们这里不考虑这个更高阶的模型．

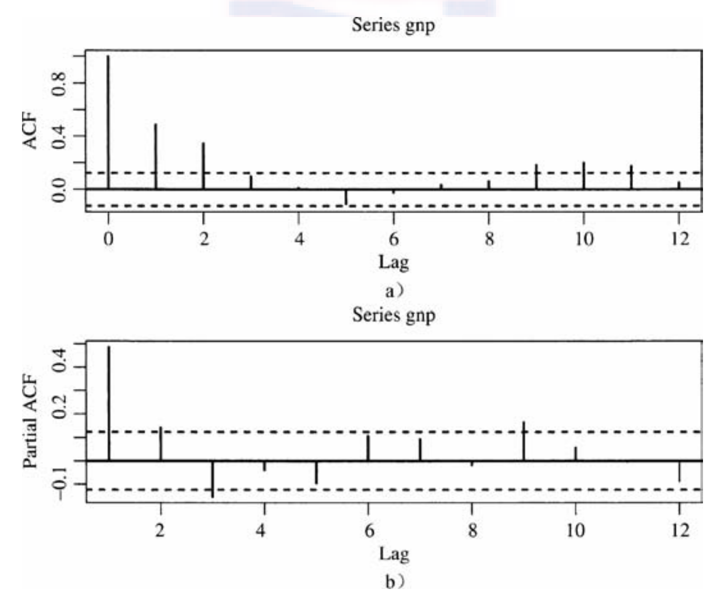


图 2-11 从 1947 年第二季度到 2010 年第一季度美国实际 GNP 季度增长率的样本偏自相关函数．虚线给出了置信水平为 95%的置信区间的近似

信息准则

有几种信息准则可用来确定 AR 过程的阶 p，它们都基于似然函数．例如，著名的赤池信息准则(Akaike Information Criterion，AIC)(Akaike，1973)的定义如下：

$$\mathrm{AIC}=\frac{-2}{T}\ln(\text{似然函数})+\frac{2}{T}(\text{参数的个数}) \tag{2-16}$$

其中 T 是样本容量，公式中的“似然函数”是指极大似然函数估计．对高斯 AR(ℓ)模型，AIC 简化为

$$\text{AIC}(\ell) = \ln(\widetilde{\sigma}_\ell^2) + \frac{2\ell}{T}$$

其中$\widetilde{\sigma}_\ell^2$ 是 σ_a^2 的极大似然估计，σ_a^2 是 a_t 的方差，T 是样本容量．式(2-16)中 AIC 的第一项度量 AR(ℓ)模型对数据的拟合优度，而第二项称为准则中的**罚函数**(penalty function，又译为补偿函数)，因为它用参数的个数来惩罚候选的模型．不同的罚函数将导致不同的信息准则．

另外一个常用的信息准则函数为施瓦茨-贝叶斯信息准则(Bayesian Information Criterion，BIC)．对高斯 AR(ℓ)模型，该准则为

$$\text{BIC}(\ell) = \ln(\widetilde{\sigma}_\ell^2) + \frac{\ell\ln(T)}{T}$$

在 AIC 中对每个参数的惩罚为 2，而在 BIC 中为 ln(T)．因此，和 AIC 相比，当样本容量适度或较大时，BIC 倾向于选择一个低阶的 AR 模型．

选择准则

在实际应用中，为了利用 AIC 来选择一个 AR 模型，要计算 AIC(ℓ)，其中ℓ=0，1，2，…，P，P 为事先给定的一个正整数，然后选择阶 k，使 AIC 达到最小值．该过程同样适用于 BIC.

表 2-1 也给出了 p=1，2，…，10 时 AIC 和 BIC 的值．这些 AIC 值都很接近，当p=9 时达到最小值－5.849，故按 AIC 准则应选 AR(9)模型．而对 BIC 来说，当 p＝1 时，达到最小值－5.833，当 p=3 时达到很接近最小值的第二小的值－5.831．因此 BIC 为加权指数收益率选择一个 AR(1)模型．这个例子说明不同的方法或者准则可能会得出 p 的不同选择．在实际应用中，还没有证据表明一种方法就一定比另外一种方法好．对给定的时间序列选择一个 AR 模型时，还有两种因素起着重要作用，就是所研究问题的具体信息和模型的简单性.

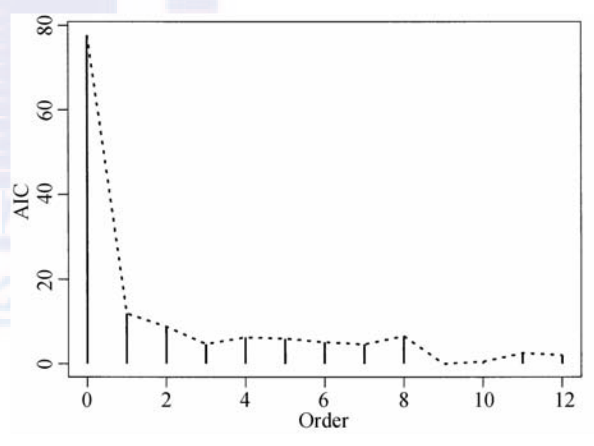


图 2-12　从 1947 年第二季度到 1991 年第一季度美国实际 GNP 季度增长率的 AIC 值．最大阶数为 P=12

61～62

再次考虑例 2.3 中美国季度实际 GNP 增长率序列．图 2-12 中给出了 P＝12 的 AIC 值．AIC 信息准则为序列识别了一个 AR(9)模型，但是它也说明了 AR(3)模型也有合理性．图 2-12 表明，如果着重于较低阶的模型，AIC 将会识别一个 AR(3)模型．注意，在 R 中的 ar 命令已经对 AIC 准则值进行了调整，使得 AIC 的最小值为 0.

```
> mm1=ar(gnp,method='mle')
> mm1$order % Find the identified order
[1] 9
> names(mm1)
 [1] "order"    "ar"         "var.pred"    "x.mean"        "aic"
 [6] "n.used"   "order.max"  "partialacf"  "resid"         "method"
[11] "series"   "frequency"  "call"        "asy.var.coef"
> print(mm1$aic,digits=3)
```

```
   0     1      2      3      4      5      6      7      8      9
77.8 11.92  8.79  4.669  6.265  5.950  5.101  4.596  6.541 0.000
    10     11     12
  0.509  2.504  2.057
> aic=mm1$aic  % For plotting below.
> length(aic)
[1] 13
> plot(c(0:12),aic,type='h',xlab='order',ylab='aic')
> lines(0:12,aic,lty=2)
```

63

参数估计

对于由式(2-9)给出的一个具体的 AR(p)模型，我们常用条件最小二乘法来估计其参数，条件最小二乘法是从第 $p+1$ 个观测值开始的．具体地说，在给定前 p 个观测值的前提下，我们有

$$x_t=\phi_0+\phi_1 x_{t-1}+\cdots+\phi_p x_{t-p}+a_t,\quad t=p+1,\cdots,T$$

上式为多元线性回归的形式，其中的参数可用最小二乘法估计．记$\hat{\phi}_i$为ϕ_i的估计，**拟合的模型**(fitted model)为

$$\hat{x}_t=\hat{\phi}_0+\hat{\phi}_1 x_{t-1}+\cdots+\hat{\phi}_p x_{t-p}$$

对应的残差为

$$\hat{a}_t=x_t-\hat{x}_t$$

称$\hat{a}_t$为**残差序列**(residual series)，并得到

$$\hat{\sigma}_a^2=\frac{\sum_{t=p+1}^{T}\hat{a}_t^2}{T-2p-1}$$

如果用条件高斯最大似然方法，ϕ_i的估计保持不变，而σ_a^2的估计变为$\tilde{\sigma}_a^2=\hat{\sigma}_a^2\times(T-2p-1)/(T-p)$．在一些统计软件包中，$\tilde{\sigma}_a^2$的定义为$\hat{\sigma}_a^2\times(T-2p-1)/T$．作为说明，考虑对表 2-1 所示的价值加权指数的月简单收益率建立一个 AR(3)模型．拟合的模型为

$$x_t=0.0091+0.116x_{t-1}-0.019x_{t-2}-0.104x_{t-3}+\hat{a}_t,\quad \hat{\sigma}_a=0.054$$

模型系数的标准误差依次分别是 0.002、0.032、0.032 和 0.032．除了间隔为 2 的系数外，其他系数在 1%的水平下都是显著的．

对这个例子，拟合模型的 AR 系数都很小，表明所考虑序列的前后相关性较弱，即使这种相关性在 1%水平下是统计显著的．拟合的模型中$\hat{\phi}_0$的显著性意味着这个序列的期望平均收益率是正的．事实上，$\tilde{\mu}=0.0091/(1-0.116+0.019+0.104)=0.009$，是一个很小的数，但它有重要的长期意义．这意味着该指数的长期收益率是相当大的．利用第 1 章中定义的多期简单收益率，这时平均年度简单毛收益率是$\left[\prod_{t=1}^{996}(1+x_t)\right]^{12/996}-1\approx0.093$．换句话说，1926—2008 年，CRSP 价值加权指数的月简单收益大约平均每年增长 9.3%，这支持这样一个通常的观点：从长期来看，股票市场的表现令人乐观．1926 年年初 1 美元的投资，到 2008 年年底的价值大约为 1593 美元．

```
> vw=read.table('m-ibm3dx.txt',header=T)[,3]
> t1=prod(vw+1)
```

64

```
> t1
[1] 1592.953
> t1^(12/996)-1
[1] 0.0929
```

模型的检验

我们必须仔细地检查拟合的模型以防止可能存在的模型的非充分性．如果模型是充分的，则其残差序列应是白噪声．残差的样本自相关函数和式(2-3)定义的 Ljung-Box 统计量可用来检验$\hat{a}_t$与白噪声的接近程度．对 AR(p)模型，Ljung-Box 统计量 $Q(m)$渐近服从自由度为 $m-g$ 的 χ^2 分布，其中 g 是所用模型中 AR 系数的个数．这里对自由度进行了修正，因为从拟合 AR(0)模型到拟合 AR(p)模型，对残差$\hat{a}_t$的限制个数增加了．如果发现拟合的模型是不充分的，那么就需要对它进行改进．例如，如果估计的 AR 系数中有一些与 0 没有显著差别，则我们应该去掉这些不显著的参数对模型进行简化．如果残差的 ACF 显示出额外的序列相关性，则应考虑到这些相关性而对模型进行扩展．

注记：大多数的时间序列软件包在对残差序列应用 Ljung-Box 统计量 $Q(m)$时不对自由度进行调整．这与理论不相符合，但是当 $m\leqslant g$ 时是可以理解的．在 R 中，在应用命令 arima 进行模型拟合后，可以应用命令 tsdiag 来进行诊断检验．具体细节后面将进行讨论． ■

现在考虑对价值加权指数的月简单收益率拟合 AR(3)模型所得的残差序列．可算得 $Q(12)=16.35$，并且基于它所渐进服从的自由度为 9 的 χ^2 分布，得到的 p 值为 0.060．这样，在 5%的水平下，前 12 个间隔无前后相关性的原假设几乎不能拒绝．然而，间隔为 2 的 AR 系数在 5%的水平下是不显著的，我们可以将模型改进为：

$$x_t = 0.0088 + 0.114x_{t-1} - 0.106x_{t-3} + a_t, \quad \hat{\sigma}_a = 0.0536$$

现在所有的估计在 1%的水平下都是显著的．残差序列给出 $Q(12)=16.83$，其 p 值为 0.078(基于 χ^2_{10} 分布)．该模型对数据的动态线性相依性的建模是充分的．

65

```
> vw=read.table('m-ibm3dx2608.txt',header=T)[,3]
> m3=arima(vw,order=c(3,0,0))
> m3
Call:
arima(x = vw, order = c(3, 0, 0))

Coefficients:
         ar1      ar2      ar3  intercept
      0.1158  -0.0187  -0.1042     0.0089
s.e.  0.0315   0.0317   0.0317     0.0017

sigma^2 estimated as 0.002875:log likelihood=1500.86, aic=-2991.7

> (1-.1158+.0187+.1042)*mean(vw)  % Compute the intercept phi(0).
[1] 0.00896761
> sqrt(m3$sigma2) % Compute standard error of residuals
[1] 0.0536189

> Box.test(m3$residuals,lag=12,type='Ljung')
        Box-Ljung test

data:  m3$residuals     % R uses 12 degrees of freedom
X-squared = 16.3525, df = 12, p-value = 0.1756
```

```
> pv=1-pchisq(16.35,9) % Compute p value using 9 degrees of freedom
> pv
[1] 0.05992276
 % To fix the AR(2) coefficient to zero:
> m3=arima(vw,order=c(3,0,0),fixed=c(NA,0,NA,NA))
 % The subcommand "fixed" enables users to fix parameter values,
 % where NA means estimation and 0 means fixing the parameter to 0.
 % The ordering of the parameters can be found using m3$coef.
> m3
Call:
arima(x = vw, order = c(3, 0, 0), fixed = c(NA, 0, NA, NA))

Coefficients:
         ar1  ar2      ar3  intercept
      0.1136    0  -0.1063     0.0089
s.e.  0.0313    0   0.0315     0.0017

sigma^2 estimated as 0.002876: log likelihood=1500.69, aic=-2993.38
> (1-.1136+.1063)*.0089  % compute phi(0)
[1] 0.00883503
> sqrt(m3$sigma2)  % compute residual standard error
[1] 0.05362832

> Box.test(m3$residuals,lag=12,type='Ljung')
        Box-Ljung test

data:  m3$residuals
X-squared = 16.8276, df = 12, p-value = 0.1562

> pv=1-pchisq(16.83,10)
> pv
[1] 0.0782113
```

66

2.4.3 拟合优度

衡量平稳模型**拟合优度**(goodness of fit)的一个常用的统计量是 R^2 统计量，其定义为

$$R^2 = 1 - \frac{\text{残差平方和}}{\text{总误差平方和}}$$

对于平稳 AR(p)时间序列模型，假设有 T 个观测值$\{x_t | t=1, \cdots, T\}$，则 R^2 变为

$$R^2 = 1 - \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{\sum_{t=p+1}^{T} (x_t - \overline{x})^2}$$

其中 $\overline{x} = \left(\sum_{t=p+1}^{T} x_t\right)/(T-p)$．容易证明，$0 \leqslant R^2 \leqslant 1$．特别地，$R^2$ 越大，表示模型对数据拟合的越好．然而该结论只对平稳时间序列成立．对于本章后面将要讨论的单位根非平稳序列，当样本容量趋于无穷大时，无论 x_t 实际服从一个什么样的模型，对其拟合的 AR(1)模型的 R^2 均趋于 1.

众所周知，对于一个给定的数据集，R^2 是所用参数个数的非降函数．为了克服该缺点，建议用**调整的 $\boldsymbol{R}^2$**(adjusted R^2)，它的定义为

$$\text{Adj}(R^2) = 1 - \frac{\text{残差的方差}}{x_t \text{ 的方差}} = 1 - \frac{\hat{\sigma}_a^2}{\hat{\sigma}_x^2}$$

其中$\hat{\sigma}_x^2$ 是 x_t 的样本方差．这个新的衡量方法将拟合模型中用到的参数个数也考虑在内．然而，其取值仍然是 0～1.

2.4.4 预测

预测是时间序列分析的一个重要应用．对式(2-9)中的 AR(p)序列，假定我们在时间指标为 h 的点上，欲预测 $x_{h+\ell}$，$\ell \geqslant 1$．时间指标 h 称为**预测原点**(forecast origin)，正整数 ℓ 称为**预测步长**(forecast horizon)．设$\hat{x}_h(\ell)$为 $x_{h+\ell}$的最小均方误差预测，我们选择$\hat{x}_h(\ell)$作为预测，该预测$\hat{x}_k(\ell)$满足：

$$E\{[x_{h+\ell} - x_h(\ell)]^2 \mid F_h\} \leqslant \min_g E[(x_{h+\ell} - g)^2 \mid F_h]$$

其中 g 是到 h 时刻(包括 h 时刻)所得到信息的函数，即 g 是 F_h 的函数．我们称$\hat{x}_k(\ell)$为 x_t 的以 h 为预测原点的 ℓ 步超前预测．在上面公式中，F_h 为在预测原点 h 所得到的信息集合.

67

超前 1 步预测

由 AR(p)模型，我们有

$$x_{h+1} = \phi_0 + \phi_1 x_h + \cdots + \phi_p x_{h+1-p} + a_{h+1}$$

在最小平方误差损失函数下，给定 F_h，x_{h+1}的点预测为条件期望

$$\hat{x}_h(1) = E(x_{h+1} \mid F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i x_{h+1-i}$$

对应的预测误差为

$$e_h(1) = x_{h+1} - \hat{x}_h(1) = a_{h+1}$$

从而，超前 1 步预测误差的方差为 $\text{Var}[e_h(1)] = \text{Var}(a_{h+1}) = \sigma_a^2$．若 a_t 服从正态分布，则 x_{h+1}的 95%的超前 1 步区间预测是$\hat{x}_h(1) \pm 1.96 \times \sigma_a$．对式(2-4)给出的线性模型，$a_{t+1}$也是以 t 为预测原点的超前 1 步预测误差．在计量经济学文献中，a_{t+1}称为在 $t+1$ 时刻对序列的扰动(shock).

在实际应用中，经常用被估参数来计算点预测和区间预测．这样做的结果是条件预测(conditional forecast)，因为这样的预测没有考虑参数估计中的不确定性．在理论上，我们可以在预测中考虑参数的不确定性，但复杂性大大增加．预测中考虑参数和模型不确定性的自然途径是马尔科夫链蒙特卡洛(Markov chain Monte Carlo，MCMC)方法．具体细节可以参见 Tsay(2010)的第 12 章．为了简单，本章假设模型为已知的．当在估计中使用的样本容量充分大时，条件预测与无条件预测是接近的.

超前 2 步预测

下面考虑以 h 为预测原点对 x_{h+2}的预测．由 AR(p)模型，我们有

$$x_{h+2} = \phi_0 + \phi_1 x_{h+1} + \cdots + \phi_p x_{h+2-p} + a_{h+2}$$

取条件期望，我们有

$$\hat{x}_h(2) = E(x_{h+2} \mid F_h) = \phi_0 + \phi_1 \hat{x}_h(1) + \phi_2 \hat{x}_h + \cdots + \phi_p x_{h+2-p}$$

对应的预测误差为

$$e_h(2) = x_{h+2} - \hat{x}_h(2) = \phi_1[x_{h+1} - \hat{x}_h(1)] + a_{h+2} = a_{h+2} + \phi_1 a_{h+1}$$

预测误差的方差为 $\mathrm{Var}[e_h(2)]=(1+\phi_1^2)\sigma_a^2$. x_{h+2}的区间预测可以用与 x_{h+1}相同的方法计算出来. 值得注意的是 $\mathrm{Var}[e_h(2)]\geqslant\mathrm{Var}[e_h(1)]$，这意味着预测步长的增加会使预测中的不确定性也增加. 这与常识是一致的：对线性序列来说，我们在 h 时刻对 x_{h+2}的把握不如对 x_{h+1}的把握大.

超前多步预测

一般地，我们有

$$x_{h+\ell} = \phi_0 + \phi_1 x_{h+\ell-1} + \cdots + \phi_p x_{h+\ell-p} + a_{h+\ell}$$

基于最小平方误差损失函数而得到的超前 ℓ 步预测就是，给定 F_h 的条件下 $x_{h+\ell}$的条件期望，可以由下式得到

$$\hat{x}_h(\ell) = \phi_0 + \sum_{i=1}^{p}\phi_i\,\hat{x}_h(\ell-i)$$

其中，当 $i\leqslant 0$ 时，容易理解$\hat{x}_h(i)=x_{h+i}$. 这个预测可用$\hat{x}_h(i)(i=1,\cdots,\ell-1)$递推地计算出来. 超前 ℓ 步预测的误差是 $e_h(\ell)=x_{h+\ell}-\hat{x}_h(\ell)$. 可以证明：对平稳 AR($p$)序列，当$\ell\to\infty$时，$\hat{x}_h(\ell)$收敛于 $E(x_t)$. 也就是说，对这样的序列，长期的点预测趋于无条件均值. 在金融学文献中，这种性质称为**均值回转**(mean reversion). 对 AR(1)模型，均值回转的速度由**半衰期**(half-life)来衡量，其中半衰期定义为 $\ell=\ln(0.5)/\ln(|\phi_1|)$. 预测误差的方差则趋于 x_t 的无条件方差. 注意，对式(2-8)中的 AR(1)模型，设 $x_t=x_t-E(x_t)$为均值调整序列. 容易得到在预测原点 h 的ℓ 步超前预测为$\hat{x}_h(\ell)=\phi_1^{\ell}x_h$. 也就是，$\phi_1^{\ell}=\frac{1}{2}$. 因此，$\ell=\ln(0.5)/\ln(|\phi_1|)$.

表 2-2 包含了对价值加权指数月简单收益率的超前 1 步至 12 步的预测结果以及相应的预测误差的标准误差：以 984 为预测原点，利用前 984 个观测值重新估计的 AR(3)模型. 拟合的模型为：

$$x_t = 0.0098 + 0.1024x_{t-1} - 0.0201x_{t-2} - 0.1090x_{t-3} + a_t$$

表 2-2 对 CRSP 价值加权指数月简单收益率用 AR(3)模型的超前多步预测的结果，预测原点为 984

步数	1	2	3	4	5	6
预测值	0.0076	0.0161	0.0118	0.0099	0.0089	0.0093
标准误差	0.0534	0.0537	0.0537	0.0540	0.0540	0.0540
真实值	−0.0623	−0.0220	−0.0105	0.0511	0.0238	−0.0786
步数	7	8	9	10	11	12
预测值	0.0095	0.0097	0.0096	0.0096	0.0096	0.0096
标准误差	0.0540	0.0540	0.0540	0.0540	0.0540	0.0540
真实值	−0.0132	0.0110	−0.0981	−0.1847	−0.0852	0.0215

注：预测原点为 984.

这里$\hat{\sigma}_a=0.054$. 表 2-2 中也给出了 2008 年的实际收益率. 由于序列的弱自相关性，预测值和预测误差的标准差很快收敛到数据的样本均值和标准差. 前 984 个观测值的样本均值和标准差分别为 0.0095 和 0.0540.

图 2-13 显示了价值加权指数月简单收益率的样本外预测值. 预测原点为 2007 年 12

月，相应于 $t=984$. 图 2-13 中包括预测值的两个标准差的上下限和 2008 年的实际收益率. 在图 2-13中，预测值和真实值分别用“o”和“·”来标记. 从图 2-13 中可知，除了 2008 年 10 月以外，所有真实收益都位于 95%的预测区间内.

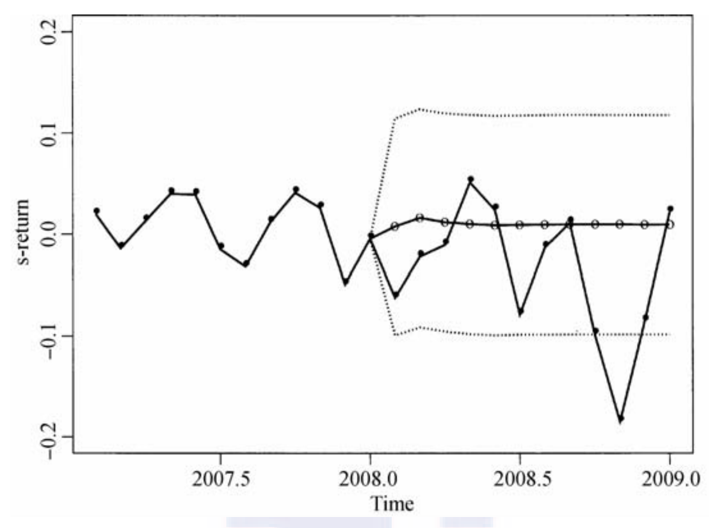


图 2-13 CRSP 价值加权指数月对数收益率的超前 1 步至 12 步的预测图，预测原点为 $t=984$，预测值用“o”表示，实际值用实点“·”表示. 两条虚线表示预测值的两个标准误差的上下限

2.5 简单移动平均模型

我们来讨论 MA 模型，它们在金融收益率建模中也很有用. 如在第 6 章将要讨论的，股票交易中买卖价的弹跳会在收益率序列中引出一个 MA(1)结构. 有几种方式引进 MA 模型：一种方式是把它当做白噪声序列的简单推广，另一种方式是把它看成参数受某种限制的无穷阶 AR 模型. 我们采用第二种方式.

69 ~ 70

除了简单性之外，没有特别的理由预先假定 AR 模型的阶是有限的. 至少在理论上我们可以考虑无穷阶 AR 模型，其形式为：

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + a_t$$

然而，这样的 AR 模型是没有实际意义的，因为它有无穷多个参数. 使这样的模型有实际用途的一个方式是假定系数 ϕ_i 满足某种限制条件，使得它们可由有限个参数决定. 这种想法的一个特殊情形为

$$x_t = \phi_0 - \theta_1 x_{t-1} - \theta_1^2 x_{t-2} - \theta_1^3 x_{t-3} - \cdots + a_t \tag{2-17}$$

其中系数只依赖于单个参数 θ_1，对 $i \geqslant 1$ 满足 $\phi_i = -\theta_1^i$. 要使式(2-17)中的模型是平稳的，θ_1 的绝对值必须小于 1；否则，θ_1^i 以及序列本身将发散. 因为 $|\theta_1| < 1$，所以当 $i \to \infty$ 时，有 $\theta_1^i \to 0$，从而 x_{t-i} 对 x_t 的贡献随 i 的增加以指数速度衰减. 这一点是合理的，因为平稳序列 x_t 对它的滞后值 x_{t-i} 如果有依赖的话，这种依赖程度应随时间衰减.

式(2-17)中的模型能写成一个比较紧凑的形式. 为此，我们把模型重写成

$$x_t + \theta_1 x_{t-1} + \theta_1^2 x_{t-2} + \cdots = \phi_0 + a_t \tag{2-18}$$

对 x_{t-1} 的模型为

$$x_{t-1}+\theta_1 x_{t-2}+\theta_1^2 x_{t-3}+\cdots=\phi_0+a_{t-1} \tag{2-19}$$

在式(2-19)两边乘以 θ_1，然后减去式(2-18)，得到

$$x_t=\phi_0(1-\theta_1)+a_t-\theta_1 a_{t-1}$$

此式说明，除了常数项外，x_t 是两个扰动 a_t 和 a_{t-1} 的加权平均. 因此，该模型称为 1 阶 MA 模型，简称 MA(1)模型. MA(1)模型的一般形式为

$$x_t=c_0+a_t-\theta_1 a_{t-1} \quad \text{或者} \quad x_t=c_0+(1-\theta_1 B)a_t \tag{2-20}$$

其中 c_0 是常数，$\{a_t\}$是白噪声序列. 类似地，MA(2)模型的形式为

$$x_t=c_0+a_t-\theta_1 a_{t-1}-\theta_2 a_{t-2} \tag{2-21}$$

MA(q)模型为

$$x_t=c_0+a_t-\theta_1 a_{t-1}-\cdots-\theta_q a_{t-q} \tag{2-22}$$

或 $x_t=c_0+(1-\theta_1 B-\cdots-\theta_q B^q)a_t$，其中 $q>0$.

71

2.5.1 MA 模型的性质

我们仍然只讨论简单的 MA(1)和 MA(2)模型，一般的 MA(q)模型的结果可用相同的方法得到.

平稳性

MA 模型总是弱平稳的，因为它们是白噪声序列的有限线性组合，其前两阶矩是不随时间变化的. 例如，考虑式(2-20)给出的 MA(1)模型. 对这个模型两端取期望，得到

$$E(x_t)=c_0$$

上式中的 $E(x_t)$不随时间变化. 在式(2-20)两端取方差，我们有

$$\mathrm{Var}(x_t)=\sigma_a^2+\theta_1^2\sigma_a^2=(1+\theta_1^2)\sigma_a^2$$

这里我们用到 a_t 与 a_{t-1} 的不相关性. $\mathrm{Var}(x_t)$也是不随时间变化的. 这些讨论对一般的 MA(q)模型也适用，因此我们得到两个一般性质：第一，MA 模型的常数项就是序列的均值(即 $E(x_t)=c_0$)；第二，MA(q)模型的方差为

$$\mathrm{Var}(x_t)=(1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2)\sigma_a^2$$

自相关函数

为简单起见，假定 MA(1)模型中 $c_0=0$. 模型两端乘以 $x_{t-\ell}$，我们有

$$x_{t-\ell}x_t=x_{t-\ell}a_t-\theta_1 x_{t-\ell}a_{t-1}$$

取期望，得到

$$\gamma_1=-\theta_1\sigma_a^2 \quad \text{且 } \ell>1 \text{ 时，} \quad \gamma_\ell=0$$

利用上述结果和 $\mathrm{Var}(x_t)=(1+\theta_1^2)\sigma_a^2$，我们有

$$\rho_0=1, \quad \rho_1=\frac{-\theta_1}{1+\theta_1^2}, \quad \rho_\ell=0, \quad \ell>1$$

因此，对 MA(1)模型，间隔为 1 的 ACF 不为 0，但所有间隔大于 1 的 ACF 都是 0. 换言之，MA(1)模型的 ACF 在间隔为 1 以后是截尾的. 对式(2-21)给出的 MA(2)模型，自相关系数是

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1+\theta_1^2+\theta_2^2}, \quad \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}, \quad \rho_\ell = 0, \quad \ell > 2 \tag{2-23}$$

这里，在间隔为 2 以后截尾．这个性质可推广到其他 MA 模型．对 MA(q)模型，其 ACF 在间隔为 q 时不为 0，但对 $\ell>q$，$\rho_\ell=0$．因此，MA(q)序列只与其前 q 个滞后值线性相关，从而它是一个“有限记忆”模型．

72

可逆性

将零均值 MA(1)模型改写为 $a_t=x_t+\theta_1 a_{t-1}$，重复代替可以得到

$$a_t = x_t + \theta_1 x_{t-1} + \theta_1^2 x_{t-2} + \theta_1^3 x_{t-3} + \cdots$$

该等式表明当前的扰动 a_t 是序列 x_t 现在和过去取值的线性组合．从直观上看，随着 j 的增加θ_1^j 应该趋于零，因为即使遥远的过去收益率 x_{t-j} 对当前的扰动有影响，其影响也微不足道．因此，要使 MA(1)模型看起来是合理的，我们应该要求$|\theta_1|<1$．这样的 MA(1)模型称为可逆的．如果$|\theta_1|=1$，则 MA(1)模型是不可逆的．对于可逆性的进一步讨论参见 Tsay(2010，第 2 章)．

2.5.2 MA 模型定阶

自相关函数 ACF 是识别一个 MA 模型阶数的有用工具．对于具有自相关函数 ρ_ℓ 的时间序列 x_t，若 $\rho_q\neq 0$，但对 $\ell>q$ 有 $\rho_\ell=0$，则 x_t 服从一个 MA(q)模型．

图 2-14 显示了 **CRSP 等权重指数**(equal-weighted index)从 1926 年 1 月到 2008 年 12 月的月简单收益率的时序图和样本自相关函数图．在样本自相关函数图中有两条虚线，它们是两个标准误差的上下限．可见，该序列的自相关系数在间隔为 1、3、9 时显著．在更大的间隔上也有稍微显著的，但我们这里不考虑它们．基于样本自相关函数，我们给该序列建立下面的 MA(9)模型：

$$x_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_3 a_{t-3} - \theta_9 a_{t-9}$$

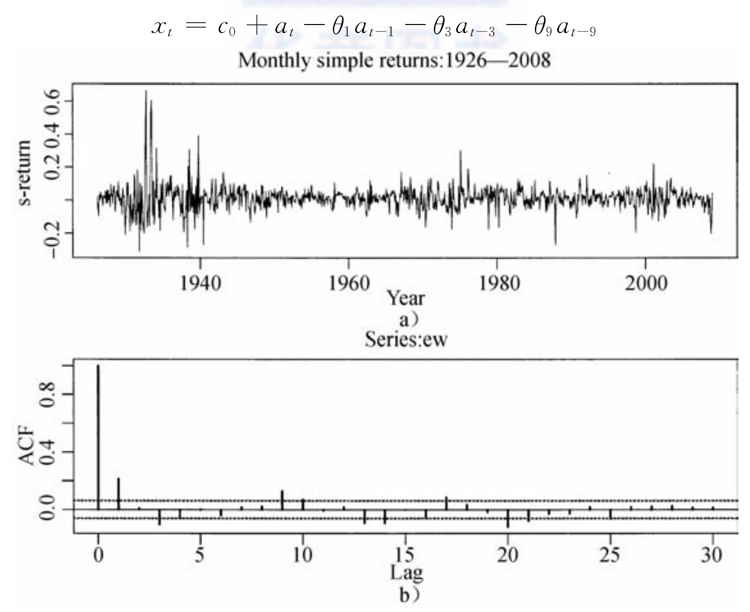


图 2-14 CRSP 等权重指数从 1926 年 1 月到 2008 年 12 月的月简单收益率的时序图和样本自相关函数图

注意，与 PACF 不同，样本 ACF 提供了模型非零 MA 滞后项的信息.

为了说明这一点，例如，考虑一个 $\theta_1=0$ 的简单 MA(2)模型. 模型为 $x_t=c_0+a_t-\theta_2 a_{t-2}$. 应用式(2-23)或者直接计算，得到模型的 ACF 为：

$$\rho_0=1,\quad \rho_1=0,\quad \rho_2=\frac{-\theta_2}{1+\theta_2^2}\text{ 并且当 } j>2 \text{ 时，有 } \rho_j=1.$$

因此对于这一特殊的例子，ACF 给出了模型结构的准确信息.

2.5.3 模型估计

估计 MA 模型通常用最大似然法. 有两种方法求 MA 模型的似然函数. 第一个方法是假设初始的“扰动”(即 a_t，$t\leqslant 0$)都是 0，这样由 $a_1=x_1-c_0$，$a_2=x_2-c_0+\theta_1 a_1$，可递推得到计算似然函数所需要的“扰动”. 这种方法称为**条件似然法**(conditional likelihood method)，所得的估计是条件最大似然估计. 第二种方法是把初始“扰动”a_t，$t\leqslant 0$，当做模型的附加参数与其他参数一起估计出来，这种方法称为**精确似然法**(exact likelihood method). 精确似然估计优于条件似然估计，尤其是当 MA 模型接近于不可逆时. 然而，精确似然估计的计算会更复杂一些. 如果样本量较大，这两种似然估计是接近的. 关于 MA 模型精确似然估计和条件似然估计的细节讨论，可参阅 Box 等(1994)或 Tsay(2010，第 8 章).

作为例子，我们来考虑 CRSP 等权重指数的月简单收益率序列，建立一个 MA(9)模型. 用条件最大似然法拟合的模型为

$$x_t=0.012+a_t+0.189a_{t-1}-0.121a_{t-3}+0.122a_{t-9},\quad \hat{\sigma}_a=0.0714 \qquad (2\text{-}24)$$

其中系数估计的标准误差分别为 0.003、0.031、0.031 和 0.031. 残差的 Ljung-Box 统计量为 $Q(12)=17.5$，基于自由度为 9 的 χ^2 分布的 p 值为 0.041. 由此可知，模型对于数据动态线性依赖关系的刻画是不充分的，还需要修正. 如果应用的自由度为 12，则 p 值为 0.132. 用精确最大似然法拟合的模型为

$$x_t=0.012+a_t+0.191a_{t-1}-0.120a_{t-3}+0.123a_{t-9},\quad \hat{\sigma}_a=0.0714 \qquad (2\text{-}25)$$

其中各估计的标准误差为 0.003、0.031、0.031 和 0.031. 残差的 Ljung-Box 统计量为 $Q(12)=17.6$. 当自由度分别为 9 和 12 时，相应的 p 值分别为 0.040 和 0.128. 和前面一样，拟合的模型仅仅略微充分. 比较模型(式(2-24))和模型(式(2-25))，可见，对这样一个具体例子来说，条件似然法和精确似然法的差别是微不足道的.

73 ~ 74

注记：R 用精确似然法进行估计. 另外，MA 的多项式写为 $1+\theta_1 B+\cdots+\theta_q B^q$，而不是通常的参数化方式 $1-\theta_1 B-\cdots-\theta_q B^q$. 特别地，R 的 `arima` 命令的 ARMA($p$，$q$)模型的形式为：

$$(1-\phi_1 B-\cdots-\phi_p B^p)(x_t-\mu)=(1+\theta_1 B+\cdots+\theta_q B^q)a_t$$

其中，μ 称为截距(intercept). 参见后面所附的 R 输出.

2.5.4 用 MA 模型预测

MA 模型的预测很容易实现. 因为模型具有有限记忆，它的点预测很快变为序列的均值. 为了说明这一点，设预测原点为 h，F_h 为在 h 时刻所能得到的信息集合. 对 MA(1)过

程的超前 1 步预测，由模型可知

$$x_{h+1} = c_0 + a_{h+1} - \theta_1 a_h$$

取条件期望，我们有

$$\hat{x}_h(1) = E(x_{h+1} | F_h) = c_0 - \theta_1 a_h$$
$$e_h(1) = x_{h+1} - \hat{x}_h(1) = a_{h+1}$$

超前 1 步预测误差的方差为 $\mathrm{Var}[e_h(1)] = \sigma_a^2$. 在实际中，$a_h$ 这个量可由多种方式得到，例如可假定 $a_0 = 0$，则有 $a_1 = x_1 - c_0$. 而当 $2 \leqslant t \leqslant h$ 时，a_t 可由公式 $a_t = x_t - c_0 + \theta_1 a_{t-1}$ 递推得到. 另外，还可用 MA(1)模型的 AR 表示来计算 a_t(参见 2.6.5 节). 当然，a_t 是拟合的 MA(1)模型的残差序列，所以它可以自然地从估计中得到.

关于超前 2 步预测，由方程

75

$$x_{h+2} = c_0 + a_{h+2} - \theta_1 a_{h+1}$$

我们有

$$\hat{x}_h(2) = E(x_{h+2} | F_h) = c_0$$
$$e_h(2) = x_{h+2} - \hat{x}_h(2) = a_{h+2} - \theta_1 a_{h+1}$$

预测误差的方差为 $\mathrm{Var}[e_h(2)] = (1+\theta_1^2)\sigma_a^2$，这是模型的方差，它大于或等于超前 1 步预测误差的方差. 上面的结果表明 MA(1)的超前 2 步预测为模型的简单无条件均值，这一点对任意预测原点 h 都正确. 更一般地，当 $\ell \geqslant 2$ 时，$\hat{x}_h(\ell) = c_0$. 总而言之，对于一个 MA(1)模型，以 h 为预测原点的超前 1 步预测为 $c_0 - \theta_1 a_h$，超前多步预测为模型的无条件均值 c_0. 如果我们画出 $\hat{x}_h(\ell)$ 相对 ℓ 变化的图形，就会看到超前 1 步预测之后的预测值变成了一条水平直线. 因此，对于 MA(1)模型而言，均值回转只需要一个时间周期.

类似地，对于 MA(2)模型，我们有

$$x_{h+\ell} = c_0 + a_{h+\ell} - \theta_1 a_{h+\ell-1} - \theta_2 a_{h+\ell-2}$$

由此得到

$$\hat{x}_h(1) = c_0 - \theta_1 a_h - \theta_2 a_{h-1}$$
$$\hat{x}_h(2) = c_0 - \theta_2 a_h$$
$$\hat{x}_h(\ell) = c_0 \quad \ell > 2$$

这样，在 2 步预测以后，MA(2)模型的超前多步预测达到序列的均值，2 步以后预测误差的方差也是序列的方差. 一般地，对一个 MA(q)模型，前 q 步以后，超前多步预测就达到模型的均值.

表 2-3 给出的是为等权重指数的月简单收益率建立的形如式(2-25)中 MA(9)模型的一些预测值，预测原点为 h=986(2008 年 2 月). 应用前 986 个观测值重新估计模型的参数. 序列观测值的样本均值和样本标准差分别是 0.0128 和 0.0736. 正如所料，表 2-3 说明：1)超前 10 步预测值即为样本均值；2)当预测步数增加时，预测误差的标准差趋于序列的标准差. 对这个具体的例子，由于次贷危机和雷曼兄弟倒闭而引起的世界范围的金融危机的影响，预测值显著地不同于观测的真实值.

表 2-3　为 CRSP 等权重指数月简单收益率拟合的 MA(9)模型的样本外预测

步数	1	2	3	4	5
预测值	0.0043	0.0136	0.0150	0.0144	0.0120
标准差	0.0712	0.0724	0.0729	0.0729	0.0729
实际值	−0.0260	0.0312	0.0322	−0.0871	−0.0010
步数	6	7	8	9	10
预测值	0.0019	0.0122	0.0056	0.0085	0.0128
标准差	0.0729	0.0729	0.0729	0.0729	0.0734
实际值	0.0141	−0.1209	−0.2060	−0.1366	0.0431

注：预测原点为 2008 年 2 月，即 $h=986$. 模型由精确极大似然法来估计.

```
> da=read.table("m-ibm3dx2608.txt",header=T)
> head(da)
      date    ibmrtn     vwrtn     ewrtn     sprtn
1 19260130 -0.010381  0.000724  0.023174  0.022472
 ....
> ew=da$ewrtn
> m1=arima(ew,order=c(0,0,9)) % unrestricted model
> m1
arima(x = ew, order = c(0, 0, 9))
Coefficients:
         ma1    ma2     ma3     ma4    ma5     ma6    ma7     ma8
      0.2144 0.0374 -0.1203 -0.0425 0.0232 -0.0302 0.0482 -0.0276
s.e. 0.0316 0.0321  0.0328  0.0336 0.0319  0.0318 0.0364  0.0354
         ma9 intercept
      0.1350    0.0122
s.e. 0.0323    0.0028
sigma^2 estimated as 0.005043: log likelihood=1220.86, aic=-2419.72
%% Refined model
> m1=arima(ew,order=c(0,0,9),fixed=c(NA,0,NA,0,0,0,0,0,NA,NA))
> m1
arima(x=ew, order=c(0,0,9), fixed=c(NA,0,NA,0,0,0,0,0,NA,NA))
Coefficients:
         ma1 ma2     ma3 ma4 ma5 ma6 ma7 ma8    ma9  intercept
      0.1909   0 -0.1199   0   0   0   0   0 0.1227     0.0122
s.e. 0.0293   0  0.0338   0   0   0   0   0 0.0312     0.0027

sigma^2 estimated as 0.005097: log likelihood=1215.61, aic=-2421.22
> sqrt(0.005097)
[1] 0.07139328
>
> Box.test(m1$residuals,lag=12,type='Ljung')  % model checking
        Box-Ljung test
data:  m1$residuals
X-squared = 17.604, df = 12, p-value = 0.1283

> pv=1-pchisq(17.6,9)  % compute p-value after adjusting the d.f.
> pv
[1] 0.04010828
%% To perform out of sample prediction at forecast origin 986.
> m1=arima(ew[1:986],order=c(0,0,9),fixed=c(NA,0,NA,0,0,0,0,0,NA,NA))
> m1
arima(x = ew[1:986], order=c(0,0,9), fixed=c(NA,0,NA,0,0,0,0,0,NA,NA))
```

```
Coefficients:
         ma1  ma2     ma3  ma4  ma5  ma6  ma7  ma8     ma9  intercept
      0.1844    0 -0.1206    0    0    0    0    0  0.1218     0.0128
s.e.  0.0295    0  0.0338    0    0    0    0    0  0.0312     0.0027

sigma^2 estimated as 0.005066: log likelihood =1206.44, aic=-2402.88
> predict(m1,10) % prediction
$pred  % point forecast
Time Series:
Start = 987
End = 996
Frequency = 1
 [1] 0.0042826 0.0135589 0.0150242 0.0144534 0.0120463 0.0018056
 [7] 0.0122115 0.0055148 0.0085135 0.0127918
$se   % standard error of prediction
Time Series:
Start = 987
End = 996
Frequency = 1
 [1] 0.071175 0.072375 0.072375 0.072882 0.072882 0.072882
 [7] 0.072882 0.072882 0.072882 0.073396
```

小结

下面依次对 AR 模型和 MA 模型进行简要总结．我们讨论了以下一些性质：

- 对 MA 模型，ACF 对模型的定阶是有用的，因为 MA(q)序列的 ACF 是 q 步截尾的．
- 对 AR 模型，PACF 对模型定阶是有用的，因为 AR(p)序列的 PACF 是 p 步截尾的．
- MA 序列总是平稳的，而对 AR 序列，当其特征根的模都小于 1 时，它才是平稳的．
- 对一个平稳序列，超前多步预测收敛于序列的均值，预测误差的方差收敛于序列的方差．

2.6 简单 ARMA 模型

在有些应用中，如果应用前面介绍的 2.4 节或者 2.5 节中的模型，可能需要高阶的 AR 或 MA 模型才能充分地描述数据的动态结构，这样就有很多参数要估计，问题就变得繁琐了．为了克服这个困难，人们提出了 ARMA 模型，参见 Box 等(1994)．ARMA 模型的基本思想是把 AR 和 MA 模型的想法结合在一个紧凑的形式中，将模型参数的个数保持很小，使模型达到参数的简约化．该模型在商业、经济和工程时间序列的建模中很有用．然而，对金融中的收益率序列，直接应用 ARMA 模型的机会较少．然而，ARMA 模型的概念与波动率建模有密切关系．事实上，广义自回归条件异方差(Generalized Autoregressive Conditional Heteroscedastic，GARCH)模型可以认为是序列 a_t^2 的 ARMA 模型，尽管是非标准的，详见第 4 章．在本节中，我们研究最简单的 ARMA(1，1)模型．

76
~
78

称一个时间序列 x_t 服从 ARMA(1，1)模型，如果 x_t 满足

$$x_t - \phi_1 x_{t-1} = \phi_0 + a_t - \theta_1 a_{t-1} \tag{2-26}$$

其中$\{a_t\}$是白噪声序列．式(2-26)的左边是模型的 AR 部分，右边是 MA 部分，常数项为 ϕ_0．为了使这样一个模型有意义，要求 $\phi_1 \neq \theta_1$；否则，在方程的两端消去一个公因子，方程所决定的过程变为一个白噪声序列．

2.6.1 ARMA(1，1)模型的性质

ARMA(1，1)模型的性质是AR(1)模型相应性质的推广，同时有一些小的修正来处理MA(1)部分的影响．首先讨论平稳性条件．在式(2-26)两端取期望，得到

$$E(x_t) - \phi_1 E(x_{t-1}) = \phi_0 + E(a_t) - \theta_1 E(a_{t-1})$$

因为对所有的 i，都有 $E(a_i)=0$，所以假定序列是弱平稳的，则 x_t 的均值为

$$E(x_t) = \mu = \frac{\phi_0}{1-\phi_1}$$

这个结果和式(2-8)的AR(1)模型的结果完全一样．

为了简单起见，假定 $\phi_0=0$．下面我们考虑 x_t 的自协方差函数．首先，在模型两端乘以 a_t 再取期望，我们有

$$E(x_t a_t) = E(a_t^2) - \theta_1 E(a_t a_{t-1}) = E(a_t^2) = \sigma_a^2 \tag{2-27}$$

把模型改写成

$$x_t = \phi_1 x_{t-1} + a_t - \theta_1 a_{t-1}$$

对上式两端取方差，得到

$$\mathrm{Var}(x_t) = \phi_1^2 \mathrm{Var}(x_{t-1}) + \sigma_a^2 + \theta_1^2\sigma_a^2 - 2\phi_1\theta_1 E(x_{t-1}a_{t-1})$$

这里我们用到 x_{t-1} 与 a_t 不相关这一事实．利用式(2-27)，我们得到

$$\mathrm{Var}(x_t) - \phi_1^2 \mathrm{Var}(x_{t-1}) = (1 - 2\phi_1\theta_1 + \theta_1^2)\sigma_a^2$$

从而，若序列 x_t 是弱平稳的，则 $\mathrm{Var}(x_t)=\mathrm{Var}(x_{t-1})$，且有

79

$$\mathrm{Var}(x_t) = \frac{(1-2\phi_1\theta_1+\theta_1^2)\sigma_a^2}{1-\phi_1^2}$$

因为方差是正的，所以我们需要 $\phi_1^2<1$(即 $|\phi_1|<1$)．这又与AR(1)模型的平稳性条件一样．

为了得到 x_t 的自协方差函数，我们假定 $\phi_0=0$，并在式(2-26)两端乘以 $x_{t-\ell}$，得到

$$x_t x_{t-\ell} - \phi_1 x_{t-1} x_{t-\ell} = a_t x_{t-\ell} - \theta_1 a_{t-1} x_{t-\ell}$$

对 $\ell=1$，在上式两端取期望并利用 $t-1$ 时的式(2-27)，我们有

$$\gamma_1 - \phi_1\gamma_0 = -\theta_1\sigma_a^2$$

其中，$\gamma_\ell=\mathrm{Cov}(x_t, x_{t-\ell})$．这个结果不同于AR(1)情形，对AR(1)模型有 $\gamma_1-\phi_1\gamma_0=0$．然而，对 $\ell=2$，取期望后，可以得到

$$\gamma_2 - \phi_1\gamma_1 = 0$$

这与AR(1)情形一样．事实上，用相同的方法可得到

$$\gamma_\ell - \phi_1\gamma_{\ell-1} = 0 \quad \ell > 1 \tag{2-28}$$

对于ACF，上述结果表明对平稳ARMA(1，1)模型，有

$$\rho_1 = \phi_1 - \frac{\theta_1\sigma_a^2}{\gamma_0} \quad \rho_\ell = \phi_1\rho_{\ell-1} \quad \ell > 1$$

这样，ARMA(1，1)模型的ACF与AR(1)模型的ACF很相似，不同之处仅在于它的指数衰减是从间隔2开始的．因此，ARMA(1，1)模型的ACF不能在任意有限间隔后截尾．

现在来看偏自相关函数(PACF)．可以证明：ARMA(1，1)模型的PACF也不能在有限间隔后截尾．它与MA(1)模型的PACF表现很相似，只是指数衰减从间隔2开始，而不

是从间隔 1 开始.

综上所述，ARMA(1，1)模型的平稳性条件与 AR(1)模型的相同，ARMA(1，1)模型的 ACF 与 AR(1)模型的 ACF 有相似的模式，只是这种相似模式是从间隔 2 开始的.

2.6.2 一般 ARMA 模型

一般 ARMA(p，q)模型的形式为

$$x_t = \phi_0 + \sum_{i=1}^{p} \phi_i x_{t-i} + a_t - \sum_{i=1}^{q} \theta_i a_{t-i}$$

其中$\{a_t\}$是白噪声序列，p 和 q 都是非负整数. AR 模型和 MA 模型都是 ARMA(p，q)的特殊情形. 利用延迟算子，上述模型可写成

80

$$(1 - \phi_1 B - \cdots - \phi_p B^p) x_t = \phi_0 + (1 - \theta_1 B - \cdots - \theta_q B^q) a_t \quad (2\text{-}29)$$

其中 $1-\phi_1 B-\cdots-\phi_p B^p$ 是模型的 AR 多项式，类似地，$1-\theta_1 B-\cdots-\theta_q B^q$ 为模型的 MA 多项式. 我们要求 AR 多项式和 MA 多项式没有公因子，否则模型的阶(p，q)会降低. 与 AR 模型一样，AR 多项式引进了 ARMA 模型的特征方程. 如果特征方程所有解的绝对值都小于 1，则该 ARMA 模型是弱平稳的. 这时，模型的无条件均值为 $E(x_t)=\phi_0/(1-\phi_1-\cdots-\phi_p)$.

2.6.3 ARMA 模型的识别

在给 ARMA 模型定阶时，ACF 和 PACF 都不能提供足够的信息. Tsay 和 Tiao(1984)提出一个新方法，它利用推广的自相关函数(Extended Autocorrelation Function，EACF)来确定 ARMA 过程的阶. EACF 的基本思想相对比较简单. 如果我们能得到 ARMA 模型的 AR 部分的相合估计，则能推导出 MA 部分. 对所推导出的 MA 序列，我们可以应用 ACF 确定 MA 部分的阶.

EACF 的推导相对复杂一些，细节参见 Tsay 和 Tiao(1984). 但此函数用起来是容易的. EACF 的结果可以用一个二维表格表示，表的行对应于 AR 的阶 p，列对应于 MA 的阶 q. ARMA(1，1)模型的 EACF 的理论形式由表 2-4 给出. 这个表的主要特征是：它包含由“O”组成的三角形，并且这个三角形左上角顶点位于阶(1，1)处. 我们正是用这样的特征来识别 ARMA 过程的阶. 一般地，对 ARMA(p，q)模型，由“O”组成的三角形的左上角顶点位于(p，q)处.

表 2-4 ARMA(1，1)模型的理论 EACF 表，其中“X”代表非零，“O”代表零，“*”代表零或者非零

	MA							
AR	0	1	2	3	4	5	6	7
0	X	X	X	X	X	X	X	X
1	X	O	O	O	O	O	O	O
2	*	X	O	O	O	O	O	O
3	*	*	X	O	O	O	O	O
4	*	*	*	X	O	O	O	O
5	*	*	*	*	X	O	O	O

注：它在识别阶(1，1)时不起任何作用.

作为例子，考虑 3M 公司股票从 1946 年 2 月到 2008 年 12 月的月对数收益率. 共有

755 个观测值，收益率序列和它的样本自相关函数如图 2-15 所示．该 ACF 图表明在 1％的水平下数据没有显著的前后相关性．表 2-5 给出了序列的样本 EACF 值和对应的一张简化表，该表由 SCA 软件包得到．也可以用 R 的添加包 TSA 来计算 EACF，具体参见下面的注记．这张简化表是用如下记号来构造的：

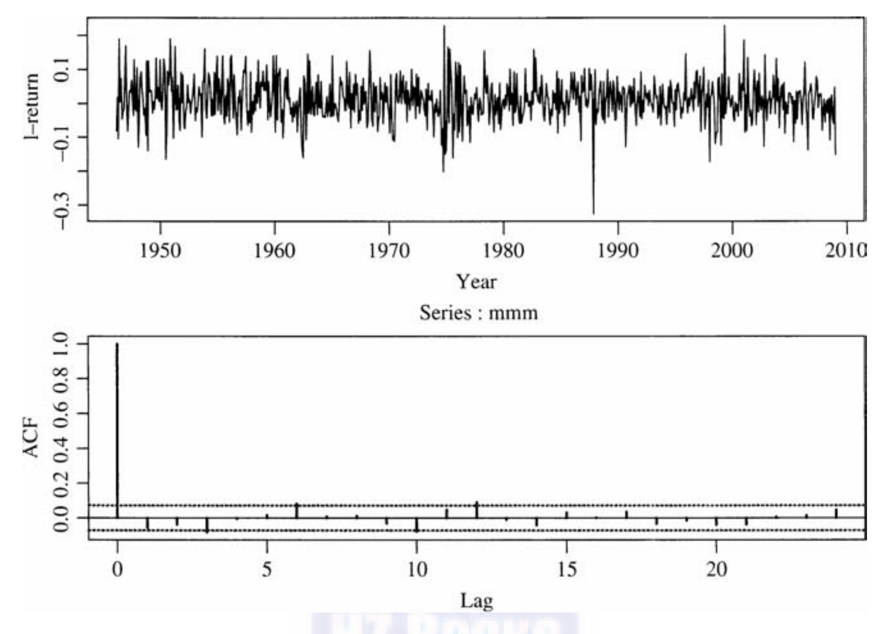


图 2-15　3M 公司股票从 1946 年 2 月到 2008 年 12 月的月对数收益率的时序图和样本自相关函数图

81～82

表 2-5　3M 公司股票从 1946 年 2 月到 1997 年 12 月的月对数收益率的样本 EACF 及其简表

a）样本推广的自相关函数

P \ MA 的阶：q	0	1	2	3	4	5	6	7	8	9	10	11	12
0	-0.06	-0.04	-0.08	-0.00	0.02	0.08	0.01	0.01	-0.03	-0.08	0.05	0.09	-0.01
1	-0.47	0.01	-0.07	-0.02	0.00	0.08	-0.03	0.00	-0.01	-0.07	0.04	0.09	-0.02
2	-0.38	-0.35	-0.07	0.02	-0.01	0.08	0.03	0.01	0.00	-0.03	0.02	0.04	0.04
3	-0.18	0.14	0.38	-0.02	0.00	0.04	-0.02	0.02	-0.00	-0.03	0.02	0.01	0.04
4	0.42	0.03	0.45	-0.01	0.00	0.00	-0.01	0.03	0.01	0.00	0.02	-0.00	0.01
5	-0.11	0.21	0.45	0.01	0.20	-0.01	-0.00	0.04	-0.01	-0.01	0.03	0.01	0.03
6	-0.21	-0.25	0.24	0.31	0.17	-0.04	-0.00	0.04	-0.01	-0.03	0.01	0.01	0.04

b）简化的 EACF 表

P \ MA 的阶：q	0	1	2	3	4	5	6	7	8	9	10	11	12
0	O	O	X	O	O	X	O	O	O	X	O	X	O
1	X	O	O	O	O	X	O	O	O	O	O	X	O
2	X	X	O	O	O	X	O	O	O	O	O	O	O
3	X	X	X	O	O	O	O	O	O	O	O	O	O
4	X	O	X	O	O	O	O	O	O	O	O	O	O
5	X	X	X	O	X	O	O	O	O	O	O	O	O
6	O	X	X	X	X	O	O	O	O	O	O	O	O

1）“X”表示对应的 EACF 的绝对值大于或等于 EACF 渐近标准误差的两倍.

2）“O”表示对应的 EACF 的绝对值小于 EACF 渐近标准误差的两倍.

可以应用 2.2 节中的 Bartlett 公式来计算 EACF 的标准误差，或者简单地用 $2/\sqrt{T}$，其中 T 为样本容量. 该简化表显示由“O”组成的三角形的左上角顶点位于阶$(p, q)=(0, 0)$处. 当 $q=2$、5、9 和 11 时，有少数“X”是例外的. 然而，该 EACF 表显示对应这些“X”的样本 ACF 值大约为 0.08 或者 0.09. 这些 ACF 值只比 $2/\sqrt{755}=0.073$ 稍大一点. 事实上，如果用 1%的临界值，简化表中的这些“X”将变为“O”. 因此，EACF 表明 3M 公司股票的月对数收益率服从一个 ARMA(0, 0)模型(也就是一个白噪声序列). 这与图 2-15 中样本 ACF 所表明的结论是一致的.

注记： 可以用 Cryer 和 Chan(2010)年开发的 R 添加包 TSA 来计算 EACF. SCA 软件包应用 Bartlett 公式来构建 EACF 简化表，而 TSA 添加包则用 $2/\sqrt{T}$作为所有 EACF 的标准误差. 对 3M 股票的收益率序列，R 的输出如下所示. ■

R 代码演示

```
> da=read.table("m-3m4608.txt",header=T)
> head(da)
      date       rtn
1 19460228 -0.077922
> mmm=log(da$rtn+1)
> library(TSA)      % Load the package
> m1=eacf(mmm,6,12)        % Simplified table
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12
0 o o x o o x o o o x o  x  o
1 x o x o o x o o o o o  x  o
2 x x x o o x o o o o o  o  o
3 x x x o o o o o o o o  o  o
4 x o x o o o o o o o o  o  o
5 x x x o x o o o o o o  o  o
6 x x x x x o o o o o o  o  o
> names(m1)
[1] "eacf"   "ar.max" "ma.ma"  "symbol"
> print(m1$eacf,digits=2)
       [,1]   [,2]   [,3]   [,4]   [,5]    [,6]    [,7]   [,8]    [,9]
[1,] -0.056 -0.038 -0.082 -0.005  0.018  0.0821  0.0080 0.0127 -0.0301
[2,] -0.474  0.010 -0.074 -0.021  0.002  0.0772 -0.0288 0.0026 -0.0068
[3,] -0.383 -0.348 -0.074  0.016 -0.006  0.0772  0.0269 0.0120  0.0004
[4,] -0.177  0.138  0.384 -0.022  0.002  0.0419 -0.0232 0.0154 -0.0044
[5,]  0.421  0.029  0.454 -0.008  0.001  0.0025 -0.0140 0.0305  0.0116
[6,] -0.114  0.214  0.449  0.010  0.202 -0.0063 -0.0038 0.0403 -0.0129
[7,] -0.208 -0.250  0.243  0.311  0.168 -0.0388 -0.0034 0.0429 -0.0101
      [,10]  [,11]   [,12]  [,13]
[1,] -0.078 0.0488  0.0909 -0.011
[2,] -0.069 0.0372  0.0938 -0.024
[3,] -0.027 0.0221  0.0428  0.042
[4,] -0.025 0.0185  0.0100  0.043
[5,]  0.004 0.0191 -0.0043  0.013
[6,] -0.012 0.0315  0.0117  0.028
[7,] -0.026 0.0078  0.0106  0.037
```

前面所讨论的信息准则也可以用来选择 ARMA 模型．具体地说，对于事先指定的正整数 P 和 Q，计算 ARMA(p，q)模型的 AIC(或 BIC)，其中 $0\leqslant p\leqslant P$、$0\leqslant q\leqslant Q$，选取使 AIC(或 BIC)取最小值的模型．该方法需要许多模型的最大似然估计，在一些情形下可能会碰到在估计时过度拟合的困难．

一旦 ARMA(p，q)模型的阶确定了，就可以用条件似然法或者精确似然法来估计模型的参数．另外，残差的 Ljung-Box 统计量可以用来检验所拟合模型的充分性．如果模型是正确的，$Q(m)$渐近地服从自由度为 $m-g$ 的χ^2 分布，其中 g 表示所用模型的 AR 或者 MA 参数的个数．

83

2.6.4 用 ARMA 模型进行预测

和 ACF 的表现类似，只要将 MA 部分对低期数预测的影响进行调整，ARMA(p，q)模型的预测与 AR(p)模型的预测有相似的特征．设预测原点为 h，F_h 为在 h 时刻所能得到的信息集合．超前 1 步预测为

$$\hat{x}_h(1)=E(x_{h+1}\mid F_h)=\phi_0+\sum_{i=1}^{p}\phi_i x_{h+1-i}-\sum_{i=1}^{q}\theta_i a_{h+1-i}$$

相应的预测误差为 $e_h(1)=x_{h+1}-\hat{x}_h(1)=a_{h+1}$．超前 1 步预测误差的方差为 $\mathrm{Var}[e_h(1)]=\sigma_a^2$．对超前 ℓ 步预测，我们有

$$\hat{x}_h(\ell)=E(x_{h+\ell}\mid F_h)=\phi_0+\sum_{i=1}^{p}\phi_i\,\hat{x}_h(\ell-i)-\sum_{i=1}^{q}\theta_i a_h(\ell-i)$$

其中，当 $\ell-i\leqslant 0$ 时，$\hat{x}_h(\ell-i)=x_{h+\ell-i}$；当 $\ell-i>0$ 时，$a_h(\ell-i)=0$；而当 $\ell-i\leqslant 0$ 时，$a_h(\ell-i)=a_{h+\ell-i}$．这样，ARMA 模型的超前多步预测可以递推算得．相应的预测误差为

$$e_h(\ell)=x_{h+\ell}-\hat{x}_h(\ell)$$

它可以用下一节给出的式(2-35)很容易地算出．

2.6.5 ARMA 模型的三种表示方式

在这一小节中，我们简单地讨论平稳 ARMA(p，q)模型的三种表示．这三种表示用于三种不同的目的．了解这三种表示可以更好地理解 ARMA 模型．第一种表示是式(2-29)给出的 ARMA(p，q)模型，这个表示很紧凑并且在参数估计时有用．另外，它在递推计算 x_t 的超前多步预测时也有用，见上一节的讨论．

对另外两种表示，我们用两个多项式的长除法．给定两个多项式 $\phi(B)=1-\sum_{i=1}^{p}\phi_i B^i$ 和 $\theta(B)=1-\sum_{i=1}^{q}\theta_i B^i$，通过长除法，我们有

$$\frac{\theta(B)}{\phi(B)}=1+\psi_1 B+\psi_2 B^2+\cdots\equiv\psi(B)\qquad(2\text{-}30)$$

和

$$\frac{\phi(B)}{\theta(B)}=1-\pi_1 B-\pi_2 B^2-\cdots\equiv\pi(B)\qquad(2\text{-}31)$$

84

例如，若 $\phi(B)=1-\phi_1 B$，$\theta(B)=1-\theta_1 B$，则有

$$\psi(B)=\frac{1-\theta_1 B}{1-\phi_1 B}=1+(\phi_1-\theta_1)B+\phi_1(\phi_1-\theta_1)B^2+\phi_1^2(\phi_1-\theta_1)B^3+\cdots$$

$$\pi(B)=\frac{1-\phi_1 B}{1-\theta_1 B}=1-(\phi_1-\theta_1)B-\theta_1(\phi_1-\theta_1)B^2-\theta_1^2(\phi_1-\theta_1)B^3-\cdots$$

由定义知 $\psi(B)\pi(B)=1$，利用 $Bc=c$ 对任意常数 c 成立这个事实(因为常数是不随时间变化的)，我们有

$$\frac{\phi_0}{\theta(1)}=\frac{\phi_0}{1-\theta_1-\cdots-\theta_q} \quad 和 \quad \frac{\phi_0}{\phi(1)}=\frac{\phi_0}{1-\phi_1-\cdots-\phi_p}$$

AR 表示

利用式(2-31)的结果，ARMA(p，q)模型可写成

$$x_t=\frac{\phi_0}{1-\theta_1-\cdots-\theta_q}+\pi_1 x_{t-1}+\pi_2 x_{t-2}+\pi_3 x_{t-3}+\cdots+a_t \tag{2-32}$$

这个表示给出了当前收益率 x_t 对过去收益率 x_{t-i} 的依赖关系，这里 $i>0$. 系数$\{\pi_i\}$称为 ARMA 模型的 **π 权重**(π-weight). 为了说明滞后值 x_{t-i}对 x_t 的贡献随 i 的增大而消失，系数 π_i 应随 i 增大而趋于零. 一个具有这样性质的 ARMA(p，q)模型，称为可逆的. 对纯 AR 模型，$\theta(B)=1$，故 $\pi(B)=\phi(B)$，这是一个有限阶的多项式. 从而对 $i>p$ 有 $\pi_i=0$，模型是可逆的. 对其他 ARMA 模型，可逆性的充分条件是：多项式 $\theta(B)$的所有零点的模大于 1. 例如，对 MA(1)模型 $x_t=(1-\theta_1 B)a_t$，一次多项式 $1-\theta_1 B$ 的零点是 $B=1/\theta_1$. 从而，如果$|1/\theta_1|>1$(即$|\theta_1|<1$)，则 MA(1)是可逆的.

由式(2-32)的 AR 表示，一个可逆的 ARMA 序列 x_t 是当前的扰动 a_t 与序列过去值的加权平均的线性组合. 对越来越遥远的过去值，权重呈指数衰减.

MA 表示

同样，利用式(2-30)长除法的结果，ARMA(p，q)模型也可以写成

$$x_t=\mu+a_t+\psi_1 a_{t-1}+\psi_2 a_{t-2}+\cdots=\mu+\psi(B)a_t \tag{2-33}$$

其中，$\mu=E(x_t)=\phi_0/(1-\phi_1-\cdots-\phi_p)$. 该表示形式清楚地说明了过去的扰动 $a_{t-i}(i>0)$对当前收益 x_t 的影响. 系数$\{\psi_i\}$称为 ARMA 模型的**脉冲响应函数**(impulse response function). 对弱平稳序列，系数 ψ_i 随 i 的增加呈指数衰减. 这一点是可以理解的，因为扰动 a_{t-i}对收益率 x_t 的影响应该随时间的流逝而消失. 这样，对平稳 ARMA 模型，扰动 a_{t-i}不能对序列有永久的影响. 如果 $\phi_0\neq 0$，那么 MA 表示中就有一个常数项，它就是 x_t 的均值(即 $\phi_0/(1-\phi_1-\cdots-\phi_p)$).

85

式(2-33)的 MA 表示在计算预测误差的方差时也有用. 在预测原点 h，我们有扰动 a_h，a_{h-1}，…. 从而，超前 ℓ 步预测为

$$\hat{x}_h(\ell)=\mu+\psi_\ell a_h+\psi_{\ell+1}a_{h-1}+\cdots \tag{2-34}$$

相应的预测误差为

$$e_h(\ell)=a_{h+\ell}+\psi_1 a_{h+\ell-1}+\cdots+\psi_{\ell-1}a_{h+1}$$

因此，超前 ℓ 步预测误差的方差为

$$\mathrm{Var}[e_h(\ell)]=(1+\psi_1^2+\cdots+\psi_{\ell-1}^2)\sigma_a^2 \tag{2-35}$$

正如所料，它是预测时间长度 ℓ 的非减函数.

最后，式(2-33)的 MA 表示还提供了平稳序列均值回转的一个简单证明．平稳性意味着当 $i\to\infty$ 时 ψ_i 趋于零．从而，由式(2-34)，我们有当 $\ell\to\infty$ 时，$\hat{x}_h(\ell)\to\mu$．因为 $\hat{x}_h(\ell)$ 是 $x_{h+\ell}$ 在预测原点 h 的条件期望，所以上述结果说的是，从长期来看，收益率序列预期会趋于它的均值，也就是说，序列是均值回转的．而且，由式(2-33)的 MA 表示，我们有 $\mathrm{Var}(x_t)=\left(1+\sum_{i=1}^{\infty}\psi_i^2\right)\sigma_a^2$．从而，由式(2-35)，当 $\ell\to\infty$ 时，我们有 $\mathrm{Var}[e_h(\ell)]\to\mathrm{Var}(x_t)$．$\hat{x}_h(\ell)$ 趋于 μ 的速度决定了均值回转的速度.

2.7 单位根非平稳性

到目前为止，我们的注意力还是集中在平稳的收益率序列上．在某些研究中，利率、汇率、资产的价格序列是研究的对象，这些序列往往是非平稳的．对于资产价格序列，其非平稳性主要是由于价格没有固定的水平．在时间序列文献中，这样的非平稳序列叫做单位根非平稳时间序列(unit-root nonstationary time serise)．单位根非平稳时间序列最著名的例子是随机游动模型.

2.7.1 随机游动

若时间序列$\{p_t\}$满足

$$p_t=p_{t-1}+a_t \tag{2-36}$$

其中 p_0 是实数，它表示该过程的起始值，而$\{a_t\}$是白噪声序列，则称$\{p_t\}$为一个**随机游动**(random walk)．若 p_t 为一只股票在第 t 天的对数价格，则 p_0 可以是该股票最初上市(initial public offering)的对数价格(即**对数 IPO 价格**)．若 a_t 的分布关于零点对称，则给定 p_{t-1} 的条件下，p_t 上升或下降的机会各有 50%，即 p_t 将随机地上升或下降．如果我们把随机游动模型看成一个特殊的 AR(1)模型，那么 p_{t-1} 的系数是 1，这不满足 AR(1)模型平稳性的条件．从而，随机游动序列不是弱平稳的，我们称之为**单位根非平稳时间序列**.

86

随机游动模型被广泛地用来作为对数股票价格运动的统计模型．在这样的模型下，股价是不可预测的或非均值回转的．为说明这一点，下面写出式(2-36)所给模型在预测原点 h 的超前 1 步预测，如下所示：

$$\hat{p}_h(1)=E(p_{h+1}\mid p_h,p_{h-1},\cdots)=p_h$$

它就是预测原点的对数股价，这样的预测没有实际意义．超前 2 步预测为

$$\begin{aligned}\hat{p}_h(2)&=E(p_{h+2}\mid p_h,p_{h-1},\cdots)=E(p_{h+1}+a_{h+2}\mid p_h,p_{h-1},\cdots)\\&=E(p_{h+1}\mid p_h,p_{h-1},\cdots)=\hat{p}_h(1)=p_h\end{aligned}$$

这又是预测原点的对数股价．事实上，对任意的预测步长 $\ell>0$，都有

$$\hat{p}_h(\ell)=p_h$$

这样，对所有预测步长，随机游动模型的点预测都是序列在预测原点的值．从而，该过程不是均值回转的.

式(2-36)给出的随机游动模型的 MA 表示为

$$p_t = a_t + a_{t-1} + a_{t-2} + \cdots$$

这个表示有几个重要的实际意义．第一，超前 ℓ 步预测误差为

$$e_h(\ell) = a_{h+\ell} + \cdots + a_{h+1}$$

从而 $\mathrm{Var}[e_h(\ell)] = \ell\sigma_a^2$，当 $\ell\to\infty$ 时，该方差发散到 ∞．因此，随着预测步长的增大，$p_{h+\ell}$ 预测区间的长度将趋于无穷．这表明，随着 ℓ 的增大，点预测 $\hat{p}_h(\ell)$ 变得没有用，这也再次说明该模型不是可预测的．第二，因为当 ℓ 增大时，$\mathrm{Var}[e_h(\ell)]$ 趋于无穷，p_t 的无条件方差是无界的．从理论上讲，这意味着对充分大的 t、p_t 可取到任何实值．对个股的对数价格来说，这一点还说得过去．但对市场指数来讲，负的对数价格即使有也是很罕见的．从这个意义上讲，随机游动模型对市场指数的适合性值得怀疑．第三，从上述表示看，对所有 i 有 $\psi_i=1$．这就是说任何过去的扰动 a_{t-i} 对 p_t 的影响不随时间衰减．从而，该序列有强记忆性，因为它记得所有过去的扰动．从经济学上讲，就是扰动对序列有持久的效应．单位根时间序列的强记忆性也可以从所观测序列的样本 ACF 看出来．当样本容量增大时，样本 ACF 都趋于 1．

87

2.7.2 带漂移的随机游动

到目前为止，前面考虑的一些实证例子中，市场指数的对数收益率序列倾向于有一个较小的正均值．这意味着对数价格的模型应为

$$p_t = \mu + p_{t-1} + a_t \tag{2-37}$$

其中 $\mu = E(p_t - p_{t-1})$，$\{a_t\}$ 是零均值的白噪声序列．模型(式(2-37))中的常数项 μ 在金融研究中是很重要的．它表示对数价格 p_t 的时间趋势，通常称为模型的**漂移**(drift)．为了说明这一点，我们假定初始价格为 p_0，从而我们有

$$\begin{aligned} p_1 &= \mu + p_0 + a_1 \\ p_2 &= \mu + p_1 + a_2 = 2\mu + p_0 + a_2 + a_1 \\ \vdots &= \vdots \\ p_t &= t\mu + p_0 + a_t + a_{t-1} + \cdots + a_1 \end{aligned}$$

上式中最后一个等式表明对数价格由时间趋势 $t\mu$ 和一个纯随机游动过程 $\sum_{i=1}^{t} a_i$ 组成．因为 $\mathrm{Var}\left(\sum_{i=1}^{t} a_i\right) = t\sigma_a^2$，其中 σ_a^2 为 a_t 的方差，p_t 的条件标准差为 $\sqrt{t}\sigma_a$，所以它增长的速度比 p_t 的条件期望慢．从而，如果我们画出 p_t 随时间指标 t 变化的图形，我们有一个斜率为 μ 的时间趋势．正斜率 μ 意味着对数价格最终趋于 ∞；相反，负斜率 μ 意味着对数价格将趋于 $-\infty$．基于这个讨论，CRSP 价值加权和等权重指数的对数收益率序列有小的、但统计显著的正均值这个现象就不足为奇了．

为了说明价格序列漂移参数的效应，我们考虑 3M 公司股票从 1946 年 2 月到 2008 年 12 月的月对数收益率．如表 2-5 中的样本 EACF 所示，该序列无显著的前后相关性．该序列服从下列简单模型

$$x_t = 0.0103 + a_t, \quad \hat{\sigma}_a = 0.0637 \tag{2-38}$$

其中 0.0103 是 x_t 的样本均值，其标准误差为 0.0023．3M 公司股票的月对数收益率在 1% 的水平下是显著不等于 0 的．事实上，应用单样本的零均值检验，t 比的值为 4.44，相应的 p 值接近于 0．我们利用对数收益率序列来构造两个对数价格序列：

88

$$p_t = \sum_{i=1}^{t} x_i \quad 和 \quad p_t^* = \sum_{i=1}^{t} a_i$$

其中 a_i 是式(2-38)中的均值修正对数收益率(即 $a_t = x_t - 0.0103$)．p_t 是 3M 公司股票的对数价格，假设初始价格为 0(即 1946 年 1 月的对数价格为 0)．p_t^* 是当对数收益率的均值为 0 时对应的对数价格．图 2-16 显示了 p_t 和 p_t^* 的时序图和直线 $y_t = 0.0103 \times t + 1946$，其中 t 是收益率序列对应的时间区间，而 1946 则是该股票的首次上市日期．从图 2-16 中可以看出，式(2-38)中常数 0.0103 的重要性是明显的．另外，如我们所料，该常数代表 p_t 上升趋势的斜率．

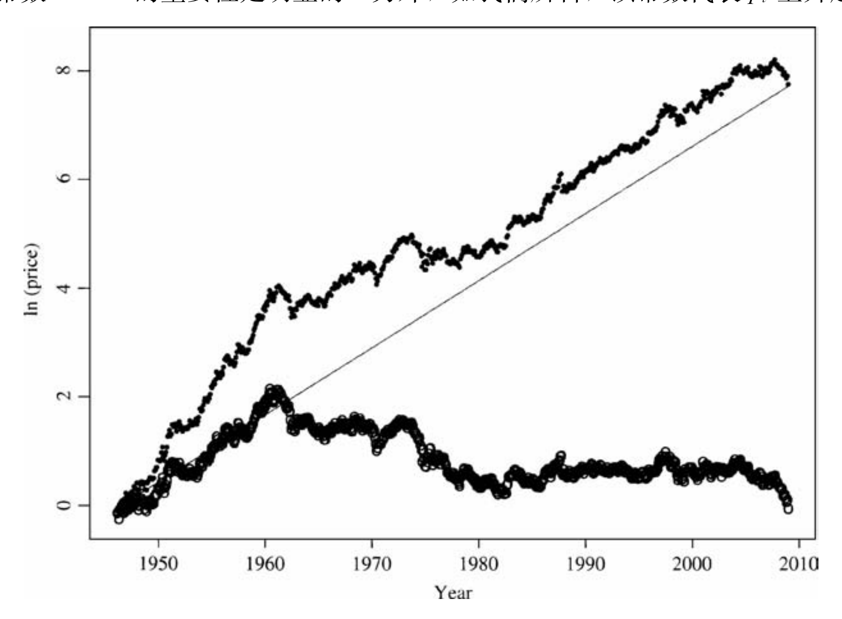


图 2-16　3M 公司股票从 1946 年 2 月到 2008 年 12 月的对数价格时序图．假定 1946 年 1 月的对数价格为 0．“o”为没有时间趋势的对数价格，直线为 $y_t = 0.0103 \times t + 1946$

常数项的解释

从前面几节的讨论中可以看出，理解时间序列模型中常数项的意义是重要的．第一，对式(2-22)中 MA(q)模型，常数项就是序列的均值；第二，对式(2-9)中的平稳 AR(p)模型或式(2-29)中的平稳 ARMA(p，q)模型，常数项与均值有关，其关系为 $\mu = \phi_0/(1-\phi_1-\cdots-\phi_p)$；第三，对带漂移的随机游动模型，常数项则变为序列的时间斜率．这些关于时间序列模型中常数项的不同解释，清楚地阐明了动态模型与通常的线性回归模型之间的不同．

动态模型与回归模型之间的另一个重要差别可以用 AR(1)模型和简单线性回归模型来说明：

$$x_t = \phi_0 + \phi_1 x_{t-1} + a_t \quad 和 \quad y_t = \beta_0 + \beta_1 x_t + a_t$$

要使 AR(1)模型有意义，其系数 ϕ_1 必须满足 $|\phi_1| \leqslant 1$. 然而，系数 β_1 可以是任何固定的实数.

2.7.3 趋势平稳时间序列

与带有线性趋势的模型紧密相关的模型是如下的趋势平稳时间序列模型：

$$p_t = \beta_0 + \beta_1 t + x_t$$

89～90

其中 x_t 是一个平稳时间序列，例如平稳 AR(p)序列. 这里，p_t 随时间以 β_1 的速率线性增长，因此它呈现出类似于带漂移的随机游动模型的行为. 然而，这两个模型有一个主要的区别. 为了说明这一点，假定 p_0 是固定的. 带漂移的随机游动模型假定均值是 $E(p_t) = p_0 + \mu t$，方差是 $\text{Var}(p_t) = t\sigma_a^2$，这两者都依赖于时间. 另一方面，趋势平稳时间序列模型假定均值是 $E(p_t) = \beta_0 + \beta_1 t$，它依赖于时间；而方差是 $\text{Var}(p_t) = \text{Var}(x_t)$，它是有限的并且不随时间变化. 趋势平稳时间序列模型可以通过简单的线性回归分析移除掉时间趋势而转换为平稳时间序列. 关于趋势平稳时间序列模型的分析，参见 2.10 节中的方法.

2.7.4 一般单位根非平稳模型

考虑 ARMA 模型. 如果我们把 ARMA 模型进行推广，允许其特征多项式以 1 作为它的特征根，则模型就变成了众所周知的**自回归求和移动平均**(Autoregressive Integrated Moving Average，ARIMA)**模型**. 因为其 AR 特征多项式有单位根，所以该 ARIMA 模型称为是单位根非平稳的. 像随机游动模型一样，ARIMA 模型有强记忆性，因为它的 MA 表示中的 ψ_i 系数不随时间衰减，这意味着过去的扰动 a_{t-i} 对序列有持久效应. 处理单位根非平稳性的惯用方法是用*差分*(differencing)方法.

差分

如果一个时间序列 y_t 变换后的序列 $c_t = y_t - y_{t-1} = (1-B)y_t$ 服从一个平稳可逆的 ARMA(p, q)模型，则称该序列为一个 ARIMA(p, 1, q)过程. 在金融中，通常认为价格序列是非平稳的，而对数收益率序列 $x_t = \ln(p_t) - \ln(p_{t-1})$ 是平稳的. 这时，对数价格序列是单位根非平稳的，从而可当做 ARIMA 过程对待. 在时间序列文献中，通过考虑时间序列相邻两值的变化量所构成的序列，把一个非平稳序列变换成一个平稳序列，这样的思想称为**差分方法**. 更形式化地说，称 $c_t = y_t - y_{t-1}$ 为 y_t 的一阶差分序列. 在某些科学领域，时间序列 y_t 可能会有多重单位根，需要做多次差分才能变成平稳序列. 例如，若 y_t 和它的一阶差分序列 $c_t = y_t - y_{t-1}$ 都是单位根非平稳的，但 $s_t = c_t - c_{t-1} = y_t - 2y_{t-1} + y_{t-2}$ 是弱平稳的，则 y_t 就有双重单位根，s_t 是 y_t 的二阶差分序列. 另外，若 s_t 服从 ARMA(p, q)模型，则 y_t 是 ARIMA(p, 2, q)过程. 对这样一个时间序列，若 s_t 有非零的均值，则 y_t 有一个二次时间函数，该二次时间函数的系数与 s_t 的均值有关. 经过季节调节后的美国季度国内生产总值(GDP)暗含的通货紧缩(implicit price deflator)序列可能有双重单位根. 然而，该序列的二阶差分序列的均值不是显著地不同于零，具体内容参见本章的习题. Box 等(1994)讨论了一般 ARIMA 模型的很多性质.

2.7.5 单位根检验

为了检验资产的对数价格 p_t 是否服从随机游动或带漂移的随机游动，我们利用如下两个模型

91

$$p_t = \phi_1 p_{t-1} + e_t \tag{2-39}$$

$$p_t = \phi_0 + \phi_1 p_{t-1} + e_t \tag{2-40}$$

其中 e_t 为误差项，考虑原假设 $H_0:\phi_1=1$ 和备择假设 $H_a:\phi_1<1$．这就是著名的单位根检验问题(Dickey 和 Fuller，1979)．一个方便的检验统计量就是在原假设下 ϕ_1 的最小二乘估计的 t 比．对式(2-39)，最小二乘法给出

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{T} p_{t-1} p_t}{\sum_{t=1}^{T} p_{t-1}^2}, \quad \hat{\sigma}_e^2 = \frac{\sum_{t=1}^{T}(p_t - \hat{\phi}_1 p_{t-1})^2}{T-1}$$

其中 $p_0=0$，T 为样本容量．t 比为

$$\mathrm{DF} \equiv t\ 比 = \frac{\hat{\phi}_1 - 1}{\hat{\phi}_1\ 的标准差} = \frac{\sum_{t=1}^{T} p_{t-1} e_t}{\hat{\sigma}_e \sqrt{\sum_{t=1}^{T} p_{t-1}^2}}$$

这个 t 比通常称为 Dickey-Fuller 检验．若$\{e_t\}$是一个白噪声序列，其稍高于二阶的矩是有限的，则当 $T\to\infty$时 DF 统计量收敛到标准布朗运动的一个函数，更多的信息参见 Chan 和 Wei(1988)以及 Phillips(1987)．如果 $\phi_0=0$，但我们采用式(2-40)，则所得的检验$\phi_1=1$ 的 t 比将趋于另一种非标准的渐近分布．在上述两种情形下，都是用模拟方法来得到检验统计量的临界值，部分临界值可参见 Fuller(1976，第 8 章)．然而，如果 $\phi_0\neq0$ 且使用的是式(2-40)，则用来检验$\phi_1=1$ 的 t 比是渐近正态的，但这需要很大的样本容量来保证渐近正态分布的成立．

对许多经济时间序列而言，ARIMA(p，d，q)可能比式(2-40)给出的简单模型更适合．在计量经济学文献中，经常使用的是 AR(p)模型．用 x_t 表示该序列，为了验证在 AR(p)过程中是否存在单位根，可以通过下列回归来检验原假设 $H_0:\beta=1$ 和备择假设 $H_1:\beta<1$

$$x_t = c_t + \beta x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t \tag{2-41}$$

其中 c_t 是关于时间指标 t 的确定性函数，$\Delta x_j = x_j - x_{j-1}$是 x_t 的差分序列．在实际中，c_t 可以是零，或者常数，或者 $c_t=\omega_0+\omega_1 t$．$\hat{\beta}-1$ 的 t 比为

$$\mathrm{ADF}\ 检验 = \frac{\hat{\beta}-1}{\hat{\beta}的标准差}$$

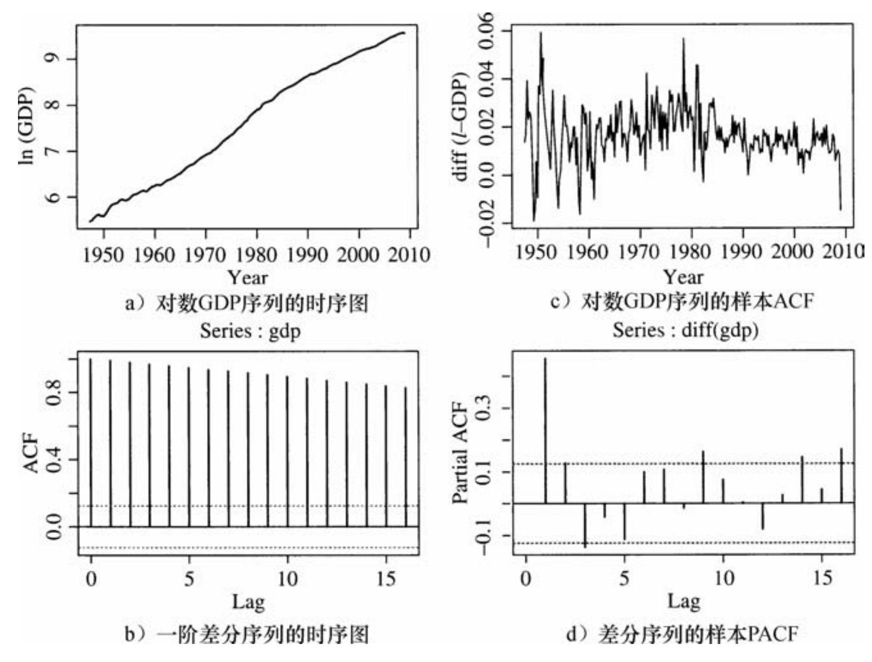
其中$\hat{\beta}$为 β 的最小二乘估计，上述 t 比就是著名的扩展 DF 单位根(Augmented Dickey-Fuller，ADF)检验．注意，由于一阶差分，式(2-41)等价于一个带确定性函数 c_t 的 AR(p)模型．式(2-41)还可以改写为

92

$$\Delta x_t = c_t + \beta_c x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t$$

其中，$\beta_c=\beta-1$．我们可以等价地检验原假设 $H_0:\beta_c=0$ 对 $H_a:\beta_c<0$．

例 2.4 考虑美国从 1947 年第 1 季度到 2008 年第 4 季度的季度 GDP 对数序列. 该序列表现出上升趋势，这表明美国经济的增长，同时它呈现高度的样本序列相关性，参见图 2-17 的左侧下方图形. 该序列的 1 阶差分序列代表了美国 GDP 的增长率，图 2-17c 给出了差分序列的时序图，该差分序列看起来在一个固定的均值水平附近波动，尽管在最近几年这种波动性比较小. 为了证实所观察到的现象，我们对该对数序列进行扩展的 Dickey-Fuller 单位根检验. 基于图 2-17 给出的差分序列的样本 PACF，我们选择 $p=10$. 我们还选用了一些其他的 p 的值，然而都没有改变检验的结论. 当 $p=10$ 时，ADF 检验统计量是 -1.611，p 值是 0.457，这表明单位根假设不能被拒绝. ■



93

图 2-17 美国从 1947 年第 1 季度到 2008 年第 4 季度的季度 GDP 对数序列

R 代码演示

```
> library(fUnitRoots)
> da=read.table("q-gdp4708.txt",header=T)
> gdp=log(da[,4])
> m1=ar(diff(gdp),method='mle')
> m1$order
[1] 10
> adfTest(gdp,lags=10,type=c("c"))
Title:
 Augmented Dickey--Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 10
  STATISTIC:
    Dickey-Fuller: -1.6109
  P-VALUE:   0.4569
```

另外一个例子，考虑标普500(S&P500)指数从1950年1月3日到2008年4月16日的日对数收益率序列，共有14 462个观测值. 图2-18给出了该序列的时序图. 如果想从实证的角度去验证该指数是否服从带漂移的随机游动，那么就有必要检验该指数序列是否有单位根. 为此，应用扩展的Dickey-Fuller检验时，我们采用$c_t=\omega_0+\omega_1 t$. 而且，基于一阶差分序列的样本PACF，我们选择$p=15$. 检验统计量的值为-1.995，p值为0.581. 因此，在任何合理的显著性水平下，单位根的原假设都不能被拒绝. 进一步的分析表明，常数项是统计显著的，尽管对时间趋势的估计不是在通常的5%的显著性水平下. 综上所述，从1950年1月到2008年4月，标普500指数的对数序列包含一个单位根和一个正的漂移项，但是没有很强的证据表明该序列有时间趋势.

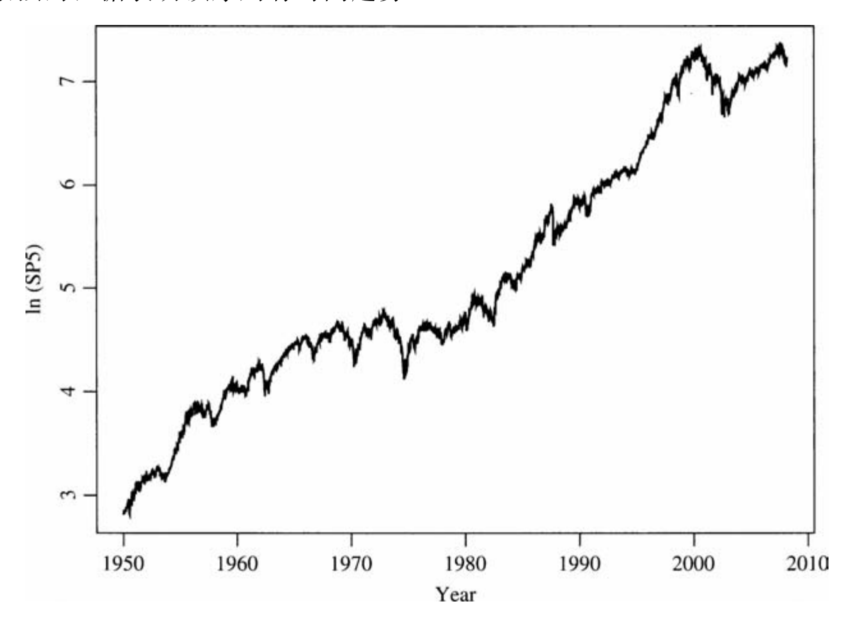


图2-18 标普500指数从1950年1月3日到2008年4月16日的对数日收益率的时序图

R代码演示

```
> library(fUnitRoots)
> da=read.table("d-sp55008.txt",header=T)
> sp5=log(da[,7])
> m2=ar(diff(sp5),method='mle') % Based on AIC
> m2$order
[1] 2
> adfTest(sp5,lags=2,type=("ct"))
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 2
  STATISTIC:
    Dickey-Fuller: -2.0179
  P-VALUE:  0.5708
```

```
> adfTest(sp5,lags=15,type=("ct")) % Based on PACF
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 15
  STATISTIC:
    Dickey-Fuller: -1.9946
  P-VALUE:   0.5807

> dsp5=diff(sp5)
> tdx=c(1:length(dsp5))
> m3=arima(dsp5,order=c(2,0,0),xreg=tdx)
> m3
Call: arima(x = dsp5,order=c(2,0,0),xreg = tdx)
Coefficients:
         ar1      ar2  intercept  tdx
      0.0721  -0.0387      4e-04    0
s.e.  0.0083   0.0083      2e-04    0
sigma^2 estimated as 8.1e-05:log likelihood=48287,aic=-96564
> m3$coef
          ar1           ar2     intercept           tdx
 7.214122e-02 -3.868823e-02  3.513995e-04 -7.165372e-09
> sqrt(diag(m3$var.coef))
         ar1          ar2     intercept          tdx
8.307510e-03 8.285669e-03 1.537309e-04 8.349685e-06
> tratio=m3$coef/sqrt(diag(m3$var.coef))
> tratio
       ar1         ar2  intercept        tdx
 8.6838549 -4.6692947  2.2858091 -0.0008582
```

2.8 指数平滑

预测中经常使用指数平滑法．我们考虑一个时间序列 x_t 的超前 1 步预测．假设预测原点是 h，所有历史数据是可知的．不失一般性，我们假设时间序列 x_t 的序列相关性以指数衰减，用历史数据的加权平均来预测 x_{h+1}，且权重的贡献是指数衰减的．我们引入一个数，使得

$$\hat{x}_{h+1} \propto wx_h + w^2 x_{h-1} + w^3 x_{h-2} + \cdots = \sum_{j=1}^{\infty} w^j x_{h+1-j}$$

其中 w 是(0，1)上的一个正实数，称为**贴现因子**(discounting rate)．然而，要想不改变 x_t 的范围，我们有必要假设权重之和为 1．利用几何序列的性质，很容易发现 $\sum_{j=1}^{\infty} w^j = \frac{1}{1-w}$．因此，一种更合适的加权平均的方式为：

$$\hat{x}_h(1) = (1-w)[wx_h + w^2 x_{h-1} + w^3 x_{h-2} + \cdots] \tag{2-42}$$

这种预测方法叫做**指数平滑法**(exponential smoothing method)．应用这种方法，数据越新对预测 x_{h+1} 的影响就越大，所以它在实际中的应用很广泛．

实际上，指数平滑法是 ARIMA 模型的特殊情形，考虑 ARIMA(0，1，1)模型：

$$(1-B)x_t = (1-\theta B)a_t$$

其中 $\theta \in (0, 1)$. 用 2.6.5 节中 AR 模型的表示方式，该模型表明

$$x_{h+1} = (1-\theta)[\theta x_h + \theta^2 x_{h-1} + \theta^3 x_{h-2} + \cdots] + a_{h+1}$$

因此，超前 1 步预测是

$$\hat{x}_h(1) = (1-\theta)[\theta x_h + \theta^2 x_{h-1} + \theta^3 x_{h-2} + \cdots]$$

94 ~ 96

令 $\theta = w$，就恰好得到式(2-42)所给出的指数平滑法.

把指数平滑法看成是 ARIMA(0，1，1)模型的特殊情形有很多优点. 首先，贴现因子 θ 可以用最大似然法估计出来；其次，可以用 ARIMA 模型的建模过程来识别和检验指数平滑法的充分性.

例 2.5 考虑芝加哥期权交易所(CBOE)从 2008 年 5 月 1 日到 2010 年 4 月 19 日的 VIX 数据. 数据来源于 CBOE 网站. 图 2-19 是对数 VIX 指数的时序图和对数 VIX 差分序列的样本 ACF. 在 5%的水平下，只有 1 阶滞后 ACF 显著不等于 0，因此可以为该差分序列建立 MA(1)模型. 令 $x_t = \ln(\text{VIX}_t)$，拟合的模型为：

$$(1-B)x_t = (1-0.163B)a_t, \quad \tilde{\sigma}_a^2 = 0.0044$$

残差的 Ljung-Box 统计量表明拟合的 ARIMA(0，1，1)模型是充分的. 比如，对于自由度为 9 的 χ^2 分布来说，我们有 $Q(10)=14.25$，$p=0.11$ 时. 因此，在本例中，我们可以用指数平滑法来预测日 VIX 指数的对数序列. ■

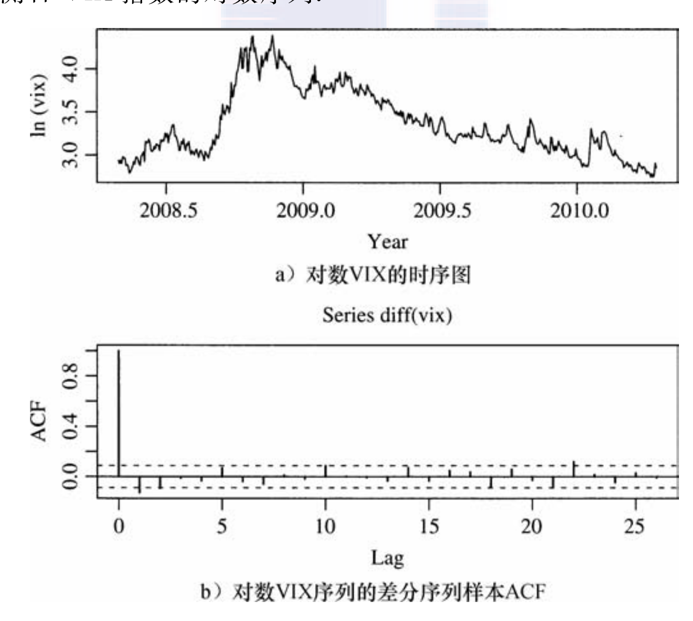


图 2-19 芝加哥期权交易所从 2008 年 5 月 1 日到 2010 年 4 月 19 日的日 VIX 指数的对数序列

97

R 代码演示

```
> da=read.table("d-vix0810.txt",header=T)
> vix=log(da$Close)
```

```
> length(vix)
[1] 496
> m1=arima(vix,order=c(0,1,1))
> m1
Call:
arima(x = vix, order = c(0, 1, 1))

Coefficients:
          ma1
      -0.1629
s.e.   0.0497

sigma^2 estimated as 0.004429: log likelihood=638.97, aic=-1273.94
>
> Box.test(m1$residuals,lag=10,type='Ljung')
        Box-Ljung test

data:  m1$residuals
X-squared = 14.2536, df = 10, p-value = 0.1617

> pp=1-pchisq(14.25,9)
> pp
[1] 0.1137060
```

最后，本节的方法可以进行推广，可以证明双指数平滑预测分析方法是 ARIMA(0，1，1)模型的一个特例。

2.9 季节模型

有些金融时间序列，如公司股票的每股季度盈利，呈现出一定的循环或周期性. 这样的时间序列叫做**季节时间序列**(seasonal time series). 回顾图 2-2 所示的可口可乐公司股票从 1983 年第 1 季度到 2009 年第 3 季度的每股季度盈利时序图. 图 2-2 中有明显的季节性，仔细检查该时序图，可以看到该季度盈利数据有很强的季节性，在样本周期内盈利呈指数上升；另外，盈利在 20 世纪 90 年代末期有一些波动. 上述的季节模式每年重复一次，从而该序列的周期是 4. 如果我们考虑月数据(如 Wal-Mart 商店的月销售额)，则周期是 12.

在与天气有关的衍生产品定价和能源期货定价方面的实证研究中，季节时间序列模型也是有用的. 众所周知，绝大部分与环境有关的时间序列显示出很强的季节性.

季节时间序列的分析已有很长的历史. 在有些应用中，季节性的重要性是次要的，可以把它从数据中移除，得到经季节调整后的时间序列，然后再用来做推断. 从时间序列中移除季节性的过程叫做**季节调整**(seasonal adjustment). 有多种季节调整的方法，例如 Peña 等(2001，第 8 章)以及它的引用文献. 美国政府公布的大多数经济数据是经季节调整的(如 GDP 增长率和失业率). 在另一些应用中，如进行预测时，数据的季节性和其他特征一样重要，必须进行相应的处理. 因为预测是金融时间序列分析的一个主要目的，所以我们着重于直接分析季节时间序列. 我们的目的是讨论一些在季节时间序列建模中有用的计量经济模型和方法.

98

2.9.1 季节差分

图 2-20 所示的是可口可乐公司的每股股票对数盈利的时序图．这里进行对数变换有两个原因：一是处理序列的指数增长．事实上，该时序图证实了取对数后的序列的确是线性增长的，这个线性增长趋势甚至在 1998 年的波动后仍然以不同的速率持续．二是对数变换可用来稳定序列的波动性．和图 2-2 比较，图 2-2 中的原始季度盈利序列有波动性增加的趋势；而在对数变换后序列的时序图中，这种增加趋势就消失了．事实上，对数变换在金融、经济时间序列分析中是常用的．在这个具体例子中，因为盈利是正的，所以在变换之前不需要做调整．而在有些情况下，我们需要对每个数据点加上一个正常数后再做对数变换．

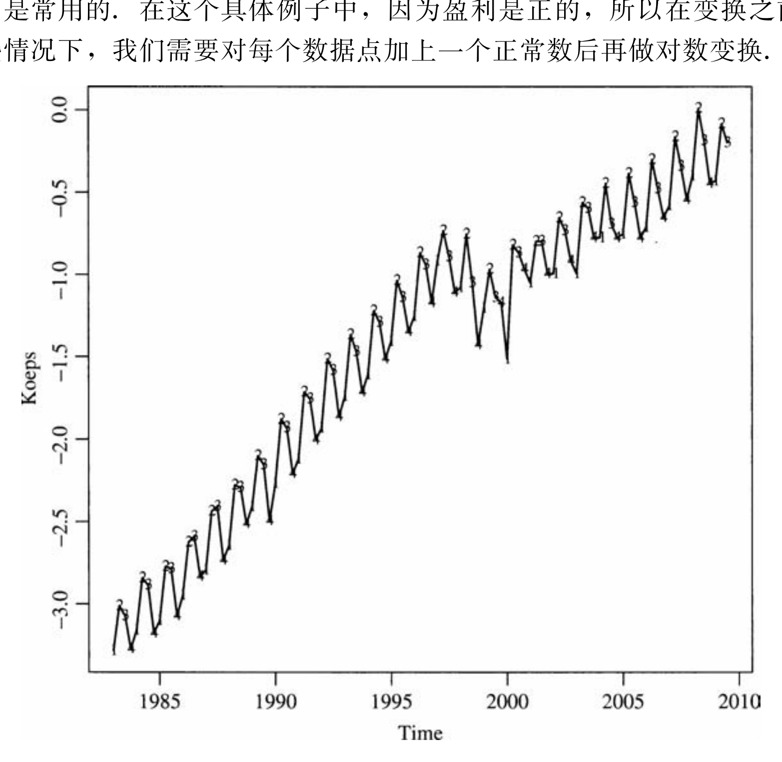


图 2-20　可口可乐公司股票从 1983 年 1 季度到 2009 年 3 季度每股季度对数盈利时序图

令对数盈利为 x_t．图 2-21a 是 x_t 的样本自相关函数图，它表明每股季度对数盈利具有强的前后相关性．处理序列强前后相关性的一个惯用方法是考虑 x_t 的一阶差分序列(即 $\Delta x_t = x_t - x_{t-1} = (1-B)x_t$)．在时间序列分析中，通过差分来得到平稳性是一种惯用手法．图 2-22a 是 Δx_t 的时序图，从图 2-22a 中可知：1)通过差分，成功地移除了数据的向上趋势；2)差分后序列显示很强的季节性．图 2-21c 给出了 Δx_t 的样本自相关函数图，可见当滞后阶数是周期 4 的倍数时自相关函数较大．另外，这种季节样本自相关函数慢慢衰减．这里我们观测到了季节时间序列的样本自相关函数的典型表现．按照 Box 等(1994，第 9 章)的内容，我们对 Δx_t 做季节差分来处理强季节模式．具体地，考虑

$$\Delta_4(\Delta x_t) = (1-B^4)\Delta x_t = \Delta x_t - \Delta x_{t-4} = x_t - x_{t-1} - x_{t-4} + x_{t-5}$$

运算 $\Delta_4 = 1 - B^4$ 叫做**季节差分**(seasonal differencing). 一般地，对一个周期为 s 的季节时间序列 y_t，季节差分意指

$$\Delta_s y_t = y_t - y_{t-s} = (1 - B^s) y_t$$

季节差分在商业和金融中常见. 例如，在报告一家公司的季度盈利时，新闻媒体通常把该盈利与该公司一年前的同一季度的盈利相比较. 通常的差分 $\Delta y_t = y_t - y_{t-1} = (1-B) y_t$ 叫做**正规差分**(regular differencing).

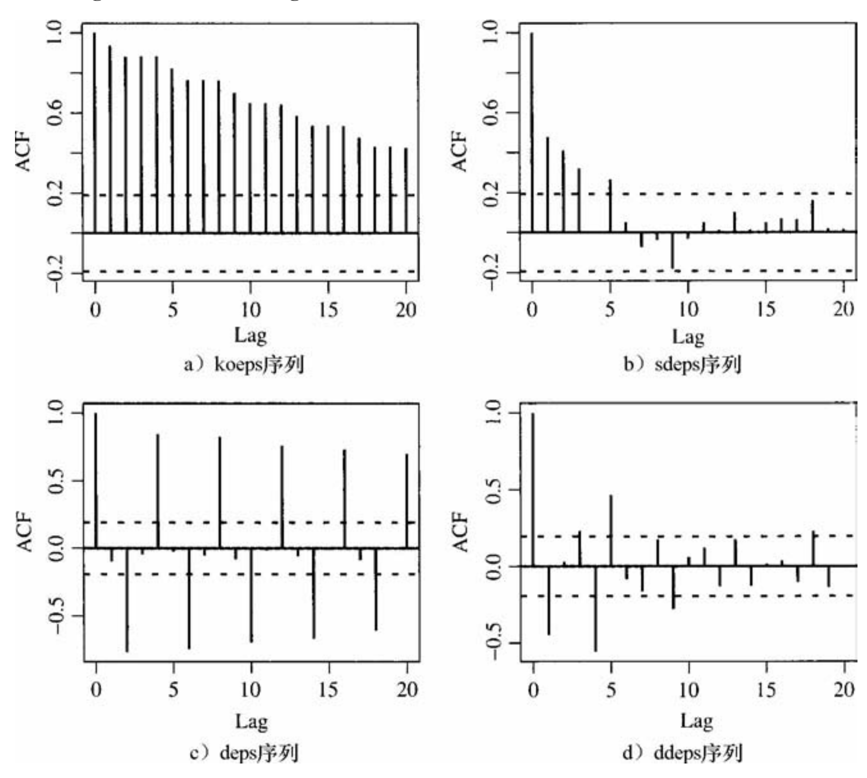


图 2-21 可口可乐公司股票从 1983 年第 1 季度到 2009 年第 3 季度每股季度盈利的对数序列的样本自相关函数图，其中“koeps”是对数盈利，“deps”是一阶差分序列，“sdeps”是季节差分序列，“ddeps”表示正规差分和季节差分后的序列

图 2-22c 给出了 $\Delta_4 \Delta x_t$ 的时序图. 从图 2-22c 中可见，序列 Δx_t 的强季节性消失了. 图 2-21d 给出了 $\Delta_4 \Delta x_t$ 的样本自相关函数 ACF 图，从该 ACF 图可以看出：

1）自相关函数在滞后阶数为 1 和 4 时统计显著且为负值. 也就是说，序列在常规滞后和季节滞后上的自相关系数不同于 0 值.

2）在滞后阶数为 5 时 ACF 为正且统计显著.

3）在滞后阶数为 3 时 ACF 为正且略微统计显著.

观测到的这些现象在实证季节时间序列中是常见的，这导致了下一节要介绍的多重季节模型的引入. 为了完整性，图 2-21 还给出了季节差分序列 $\Delta_4 x_t$ 的样本自相关函数图，图 2-22b 则给出了相应的时序图.

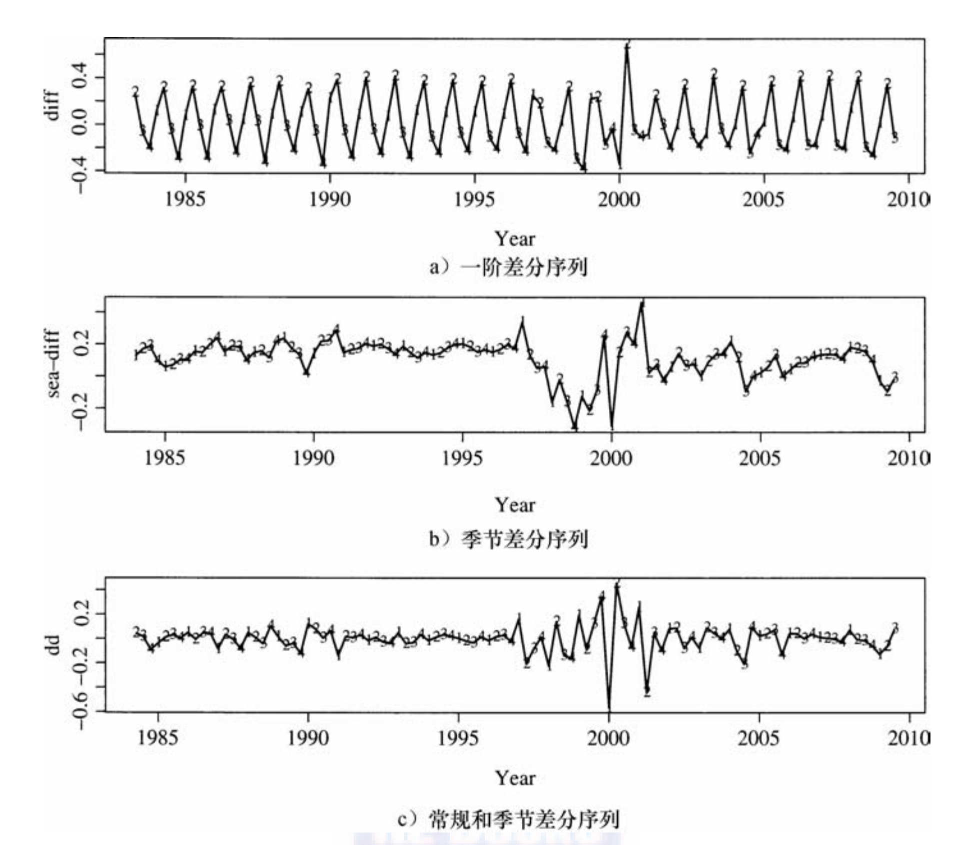


图 2-22 可口可乐公司股票从 1983 年第 1 季度到 2009 年第 3 季度每股季度盈利的对数序列的时序图

2.9.2 多重季节模型

如果一个统计模型的自相关函数具有图 2-21 中 $(1-B^4)(1-B)x_t$ 的样本 ACF 的表现，则该模型称为多重季节模型，即

$$(1-B)(1-B^s)x_t = (1-\theta B)(1-\Theta B^s)a_t \tag{2-43}$$

其中 s 是序列的周期，a_t 是白噪声序列，$|\theta|<1$，$|\Theta|<1$. 此模型在文献中称为**航空模型**（airline model），见 Box 等（1994，第 9 章）. 它广泛地应用于季节时间序列的建模. 此模型的 AR 部分由正规差分和季节差分两部分组成，而 MA 部分包括两个参数. 下面着重讨论该模型的 MA 部分：

$$w_t = (1-\theta B)(1-\Theta B^s)a_t = a_t - \theta a_{t-1} - \Theta a_{t-s} + \theta\Theta a_{t-s-1}$$

其中 $w_t=(1-B^s)(1-B)x_t$，且 $s>1$. 容易得到 $E(w_t)=0$，并且

$$\mathrm{Var}(w_t) = (1+\theta^2)(1+\Theta^2)\sigma_a^2$$

$$\mathrm{Cov}(w_t, w_{t-1}) = -\theta(1+\Theta^2)\sigma_a^2$$

$$\mathrm{Cov}(w_t, w_{t-s+1}) = \theta\Theta\sigma_a^2$$

$$\mathrm{Cov}(w_t, w_{t-s}) = -\Theta(1+\theta^2)\sigma_a^2$$
$$\mathrm{Cov}(w_t, w_{t-s-1}) = \theta\Theta\sigma_a^2$$
$$\mathrm{Cov}(w_t, w_{t-\ell}) = 0, \quad 其中\ \ell \neq 0, 1, s-1, s, s+1$$

因此，序列 w_t 的 ACF 为

$$\rho_1 = \frac{-\theta}{1+\theta^2}, \quad \rho_s = \frac{-\Theta}{1+\Theta^2}, \quad \rho_{s-1} = \rho_{s+1} = \rho_1\rho_s = \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)}$$

99 ~ 102

对 $\ell>0$ 且 $\ell\neq1$，$s-1$，s，$s+1$，有 $\rho_\ell=0$. 例如，如果 w_t 是季度时间序列，那么 $s=4$ 且它的 ACF 只在滞后阶数为 1、3、4、5 时非零. 事实上，这就是可口可乐公司股票季度对数盈利序列的情况.

把上述 ACF 与 MA(1)模型 $y_t=(1-\theta B)a_t$ 和 MA(s)模型 $z_t=(1-\Theta B^s)a_t$ 的 ACF 相比较，会看到有意思的结论. y_t 和 z_t 序列的 ACF 分别是

$$\rho_1(y) = \frac{-\theta}{1+\theta^2}, \quad 且\ \rho_\ell(y) = 0, \quad \ell > 1$$

$$\rho_s(z) = \frac{-\Theta}{1+\Theta^2}, \quad 且\ \rho_\ell(z) = 0, \quad \ell > 0, \quad \ell \neq s$$

我们看到：1)$\rho_1=\rho_1(y)$；2)$\rho_s=\rho_s(z)$；3)$\rho_{s-1}=\rho_{s+1}=\rho_1(y)\times\rho_s(z)$. 因此，序列 w_t 在滞后阶数为 $s-1$ 和 $s+1$ 时 ACF 可以认为是滞后阶数为 1 和间隔为 s 的序列相关性的*交互作用*(interaction)的结果. w_t 的模型称为**多重**(multiplicative)**季节 MA 模型**. 在实际中，多重季节模型意味着序列的正规部分与季节部分的动态结构是近似正交的.

为说明航空模型的有用性，重新改写式(2-43)所表示的航空模型为下式：

$$\frac{1-B}{1-\theta B}\left(\frac{1-B^s}{1-\Theta B^s}x_t\right) = a_t$$

设 $y_t=(1-B^s)/(1-\Theta B^s)x_t$，那么有

$$(1-B)y_t = (1-\theta B)a_t, \quad (1-B^s)x_t = (1-\Theta B^s)y_t$$

这里，y_t 是 2.8 节中的指数平滑模型，而 x_t 为另一个带有季节成分的指数平滑模型. 所以，航空模型可以认为是作用于另一个指数平滑模型的指数平滑模型. 其中的一个指数平滑用于通常的序列相关性，而另一个指数平滑用于季节相关性.

例 2.6 在本例中，我们把航空模型应用到可口可乐公司股票从 1983 年 1 季度到 2009 年 3 季度每股季度盈利的对数序列. 基于精确似然法所拟合的模型为

$$(1-B)(1-B^4)x_t = (1-0.4096B)(1-0.8203B^4)a_t, \quad \hat{\sigma}_a^2 = 0.007\,24$$

其中两个 MA 参数的标准误差分别为 0.0866 和 0.0743，残差的 Ljung-Box 统计量为 $Q(12)=13.2$，当自由度为 12 时，其 p 值为 0.35. 当调整自由度为 10 时，相应的 p 值为 0.21. 图 2-23 给出了拟合的航空模型的诊断图. 图 2-23a 是标准化残差，该图可以用于检查残差的 iid 假设，并找出数据中可能存在的异常值. 图 2-23b 给出了残差的 ACF 图. 理想情况下，所有的残差 ACF 应该在两倍的标准误差上下限内. 图 2-23c 给出了几个不同 m 值对应的 Ljung-Box 统计量的 p 值. 如果拟合的模型能充分描述数据中的序列相关性，那

么所有这些 p 值应该大于第Ⅰ类错误值. 图 2-23 中的虚线给出了默认的第Ⅰ类错误值 0.05. m 的具体数值可以通过命令 tsdiag 的子命令 gof 来指定. 基于图 2-23 给出的 3 幅图形，除了在 1999 年年末可能有异常值外，模型看起来充分拟合了可口可乐公司股票的季度对数盈利数据. ■

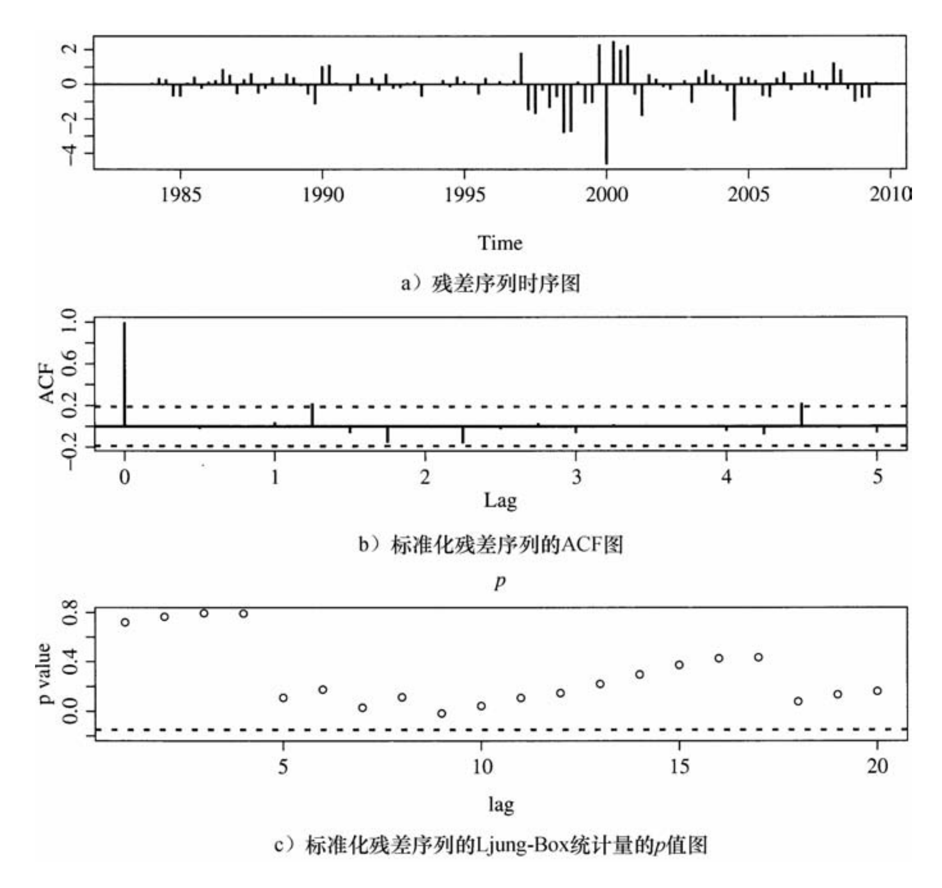


图 2-23 可口可乐公司股票从 1983 年第 1 季度到 2009 年第 3 季度每股季度盈利对数序列的拟合的航空模型诊断图

为了说明上述拟合的季节模型的预测性能，我们利用前 100 个观测值，即从 1983 年到 2007 年的数据，重新估计模型的参数，而把最后 7 个数据点用来进行预测评价. 重新拟合的模型为：$(1-B)(1-B^4)x_t=(1-0.4209B)(1-0.8099B^4)a_t$，$\hat{\sigma}_a^2=0.00743$，以 $h=100$ 为预测原点，计算超前 1～7 步预测值和它们的标准误差. 为了得到每股盈利的预测值，利用第 1 章中给出的正态分布与对数正态分布之间的关系，先进行反对数变换. 图 2-24 给出了模型的预测表现，其中实线表示原始序列数据，预测期的真实观测值用“o”标识，点预测值用“*”标识. 虚线表示 95％的区间预测. 从图 2-24 中可看出，预测显示出强烈的季节性，并与实际观测值相近. 真实的盈利都位于区间预测内.

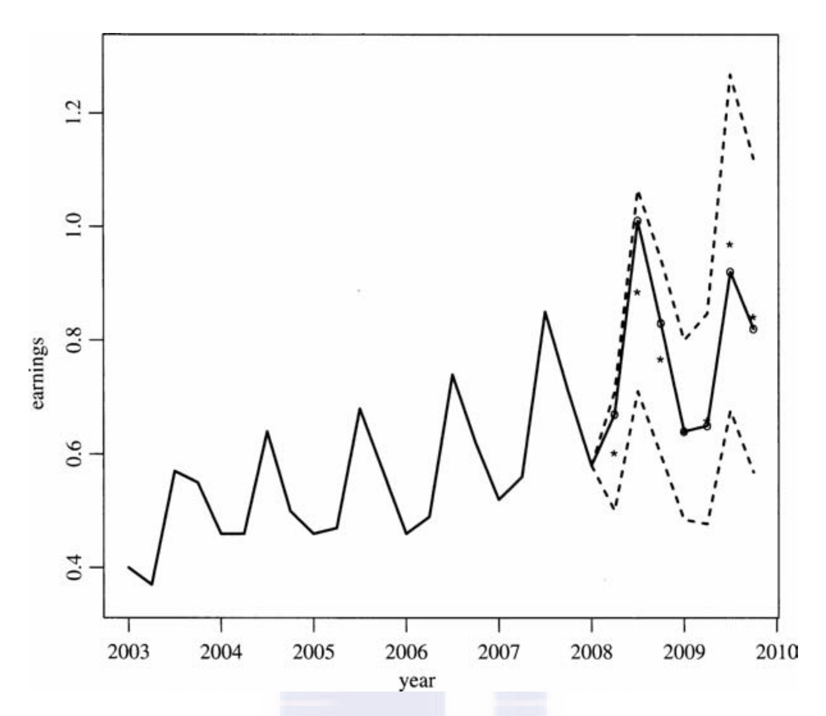


图 2-24 可口可乐公司股票季度盈利的样本外点预测和区间预测. 预测原点为 2007 年. 预测期的真实观测值用“o”标识，点预测值用“ * ”标识，虚线表示 95％的区间预测

R 代码演示

```
> da=read.table("q-ko-earns8309.txt",header=T)
> head(da)
     pends   anntime  value
1 19830331 19830426 0.0375
...
6 19840630 19840720 0.0583
> eps=log(da$value)
> koeps=ts(eps,frequency=4,start=c(1983,1))
> plot(koeps,type='l')
> points(koeps,pch=c1,cex=0.6)
% Obtain ACF plot
> par(mfcol=c(2,2))
> koeps=log(da$value)
> deps=diff(koeps)
> sdeps=diff(koeps,4)
> ddeps=diff(sdeps)
> acf(koeps,lag=20)
> acf(deps,lag=20)
> acf(sdeps,lag=20)
> acf(ddeps,lag=20)
% Obtain time plots
> c1=c("2","3","4","1")
> c2=c("1","2","3","4")
> par(mfcol=c(3,1))
```

```
> plot(deps,xlab='year',ylab='diff',type='l')
> points(deps,pch=c1,cex=0.7)
> plot(sdeps,xlab='year',ylab='sea-diff',type='l')
> points(sdeps,pch=c2,cex=0.7)
> plot(ddeps,xlab='year',ylab='dd',type='l')
> points(ddeps,pch=c1,cex=0.7)
%  Estimation
> m1=arima(koeps,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=4))
> m1
Call:
arima(x=koeps,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=4))
Coefficients:
          ma1      sma1
      -0.4096   -0.8203
s.e.   0.0866    0.0743

sigma^2 estimated as 0.00724:  log likelihood = 104.25,  aic = -202.5

> tsdiag(m1,gof=20)  % model checking
> Box.test(m1$residuals,lag=12,type='Ljung')
        Box-Ljung test
data:  m1$residuals
X-squared = 13.3034, df = 12, p-value = 0.3474
> pp=1-pchisq(13.30,10)
> pp
[1] 0.2073788    % p-value
% Out-of-sample forecasting
> koeps=log(da$value)
> length(koeps)
[1] 107
> y=koeps[1:100]
> m1=arima(y,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=4))
> m1
Coefficients:
          ma1      sma1
      -0.4209   -0.8099
s.e.   0.0874    0.0767

sigma^2 estimated as 0.007432:  log likelihood = 95.78,  aic = -185.57
% Prediction
> pm1=predict(m1,7)
> names(pm1)
[1] "pred" "se"
> pred=pm1$pred
> se=pm1$se
% Anti-log transformation
> ko=da$value
> fore=exp(pred+se^2/2)
> v1=exp(2*pred+se^2)*(exp(se^2)-1)
> s1=sqrt(v1)
> eps=ko[80:107]
> length(eps)
[1] 28
> tdx=(c(1:28)+3)/4+2002
> upp=c(ko[100],fore+2*s1)
> low=c(ko[100],fore-2*s1)
```

```
> min(low,eps)
[1] 0.37
> max(upp,eps)
[1] 1.267623
> plot(tdx,eps,xlab='year',ylab='earnings',type='l',ylim=c(0.35,1.3))
> points(tdx[22:28],fore,pch='*')
> lines(tdx[21:28],upp,lty=2)
> lines(tdx[21:28],low,lty=2)
> points(tdx[22:28],ko[101:107],pch='o',cex=0.7)
```

在某些应用中，需要应用如下的多重季节模型：

$$(1-B)(1-B^s)x_t=(1-\theta_1 B-\theta_2 B^2)(1-\Theta B^s)a_t$$

特别是，当 $s>4$ 时．这里，差分序列 $w_t=(1-B)(1-B^s)x_t$ 的 ACF 可能在滞后阶数为 1、2、$s-2$、$s-1$、s、$s+1$ 和 $s+2$ 时非零．w_t 的样本 ACF 可以用来识别这样的模型．

模型

$$w_t=(1-\theta B-\Theta B^s)a_t \tag{2-44}$$

其中，$|\theta|<1$ 和 $|\Theta|<1$，是一个非乘积季节 MA 模型．显而易见，在式(2-44)的模型中有 $\rho_{s+1}=0$．乘积模型比对应的非乘积模型更节省，因为虽然两个模型都用了相同数目的参数，但是乘积模型有更多非零的 ACF．

2.9.3 季节哑变量

当一个时间序列的季节模式是随时间稳定(即，近似于确定性函数)时，可用哑变量(dummy variable)来处理季节性．这里所谓的季节哑变量，是指用指示变量来代表一年中的各个季节．对于季度数据，哑变量分别代表春季、夏季、秋季和冬季，其中的 3 个哑变量将用于分析中．许多分析师采用这种分析方法．然而，确定性的季节只是前面讨论的多重季节模型的一个特殊情形．特别地，如果 $\Theta=1$ 时，模型(式(2-43))包含一个确定性的季节成分．因此，当季节模式是确定性时，用哑变量或者多重季节模型都能得到相同的预测结果．但是，当季节模式不是确定性的时候，哑变量方法会导致较差的预测．在实际中，我们建议用精确似然法去估计一个多重季节模型，特别是当样本容量较小或可能存在一个确定性的季节成分时．

103～107

例 2.7 为了说明确定性的季节性，考虑 CRSP 第 1 个 10 分位(市值最大的公司)指数从 1970 年 1 月到 2008 年 12 月的月简单收益率，共有 468 个观测值．图 2-25a 给出了序列的时序图，该时序图并没有显示出序列有明显的季节性．然而，图 2-25b 给出的样本 ACF 在滞后阶数为 12、24、36 以及滞后阶数为 1 处显著不为零．如果选用季节 ARMA 模型，则模型具有以下形式：

$$(1-\phi_1 B)(1-\phi_{12}B^{12})X_t=(1-\theta_{12}B^{12})a_t$$

其中 X_t 表示月简单收益率．去除不显著的参数后，拟合的模型为：

$$(1-0.179B)(1-0.989B^{12})X_t=(1-0.913B^{12})a_t,\quad \tilde{\sigma}_a^2=0.00472$$

从拟合的模型中可以清楚地看到，季节 AR 和 MA 因子近乎相互抵消．这表明精确似然方法应用的有效性，并且估计的结果说明可能存在确定性的季节性．为了进一步证实确定性

的季节性存在的论断，我们定义关于 1 月的哑变量，即

$$\text{Jan}_t = \begin{cases} 1 & t \text{ 是一月} \\ 0 & \text{其他月} \end{cases}$$

并应用简单线性回归

$$X_t = \beta_0 + \beta_1 \text{Jan}_t + e_t$$

拟合的模型为 $X_t = 0.0029 + 0.1253\text{Jan}_t + e_t$，其中估计的标准误差分别为 0.0033 和 0.0115. 图 2-25a、b 分别给出了上述简单线性回归残差的时序图和残差的样本 ACF. 从样本 ACF 来看，在滞后阶数为 12、24 和 36 处的序列相关性大大消除了. 这表明通过 1 月哑变量，已经成功地消除了第 1 个 10 分位收益率的季节模式. 因此，第 1 个 10 分位的月简单收益率的季节性主要是由于 **1 月效应**(January effect)导致的. ■

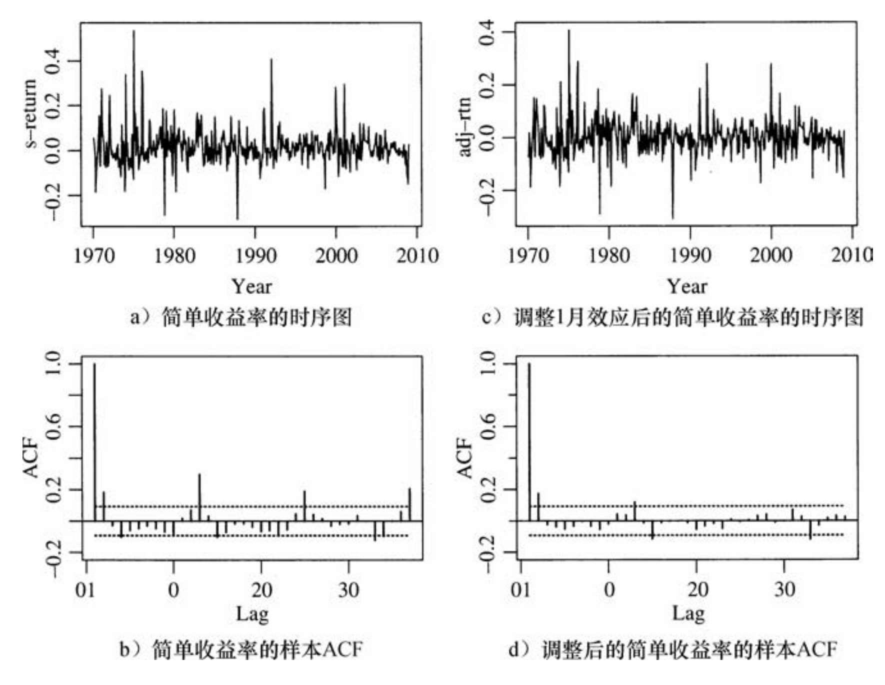


图 2-25　CRSP 第 1 个 10 分位指数从 1970 年 1 月到 2008 年 12 月的月简单收益率

R 代码演示(输出经过编辑，%表示注解)

```
> da=read.table("m-deciles08.txt",header=T)
> d1=da[,2]
> jan=rep(c(1,rep(0,11)),39) % Create January dummy.
> m1=lm(d1 ~ jan)
> summary(m1)
lm(formula = d1 ~ jan)

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.002864    0.003333   0.859    0.391
jan         0.125251    0.011546  10.848   <2e-16 ***
```

```
---

Residual standard error: 0.06904 on 466 degrees of freedom
Multiple R-squared: 0.2016,     Adjusted R-squared: 0.1999

> m2=arima(d1,order=c(1,0,0),seasonal=list(order=c(1,0,1),
+ period=12))
> m2
Coefficients:
         ar1     sar1      sma1   intercept
      0.1769   0.9882   -0.9144     0.0118  % insignificant intercept
s.e.  0.0456   0.0093    0.0335     0.0129

sigma^2 estimated as 0.004717: log likelihood=584.07, aic=-1158.14
> tsdiag(m2,gof=36)  % plot not shown.

> m2=arima(d1,order=c(1,0,0),seasonal=list(order=c(1,0,1),
+ period=12),include.mean=F)
> m2

Call:
arima(x=d1,order=c(1,0,0),seasonal=list(order=c(1,0,1),period=12),
    include.mean = F)

Coefficients:
         ar1     sar1      sma1
      0.1787   0.9886   -0.9127   % All estimates are significant.
s.e.  0.0456   0.0089    0.0335

sigma^2 estimated as 0.00472: log likelihood=583.68, aic=-1159.36
```

2.10 带时间序列误差的回归模型

在许多应用中，主要兴趣在于两个时间序列的关系上．金融中的**市场模型**(Market Model)显然就是其中一个例子，它要找出个股的超额收益率与市场指数收益率之间的关系．利率的期限结构是另一个例子，它研究的是不同期限利率之间的关系是怎样随时间演变的．这些例子导致我们考虑如下形式的线性回归：

$$y_t = \alpha + \beta x_t + e_t \tag{2-45}$$

其中 y_t 和 x_t 是两个时间序列，e_t 表示误差项．经常用最小二乘(LS)法来估计模型(式(2-45))．若$\{e_t\}$是白噪声序列，则 LS 给出的估计是相合的．然而，在实际中经常遇到误差 e_t 是前后相关的情形．这时，模型(式(2-45))就是一个带时间序列误差的回归模型，最小二乘法(LS)所产生的 α 和 β 的估计可能是不相合的．

带时间序列误差的回归模型在经济、金融中有广泛应用．然而，应用该模型必须小心，不要忽视 e_t 的前后相关性．此模型值得我们仔细研究．

我们通过考虑如下两个美国周利率序列之间的关系来介绍该模型：

1) x_{1t}：一年期固定期限国库券利率．

2) x_{3t}：三年期固定期限国库券利率．

这两个序列都是以百分数给出，观测时间都是从 1962 年 1 月 5 日至 2009 年 10 月 4 日，都有 2467 个观测值，数据是从圣路易斯联邦储备银行得到的．严格地讲，应该利用多元时间序列分析对这两个利率序列联合建模（Tsay，2010，第 8 章）．然而，为了简单起见，我们将注意力放在回归类型的分析上，并且忽略它们的同时性．

图 2-26 是上述两个利率序列的时序图，实线是一年期的利率，虚线是三年期的利率．图 2-27a 是 x_{1t}和 x_{3t}的散点图，正好与预期的一样，这两种利率是高度相关的．描述这两种利率之间关系的一个简单方式是利用简单模型 $x_{3t}=\alpha+\beta x_{1t}+e_t$，所拟合的模型是

108～110

$$x_{3t}=0.832+0.930x_{1t}+e_t,\quad \hat{\sigma}_e=0.523 \tag{2-46}$$

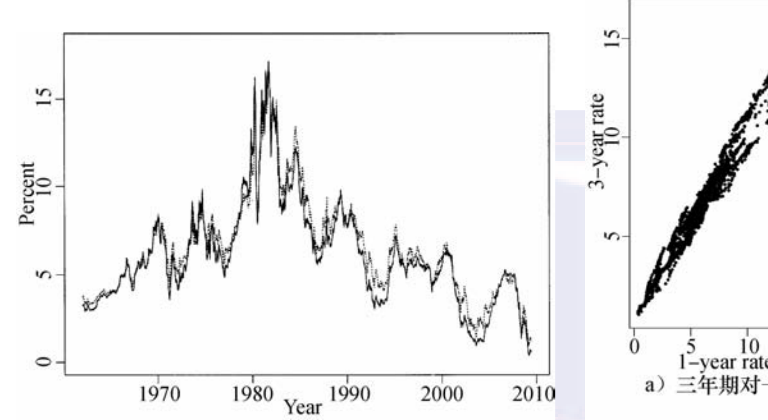


图 2-26　从 1962 年 1 月 5 日到 2009 年 4 月 10 日美国周利率（百分比）的时序图：实线是一年期固定期限国债利率，虚线是三年期固定期限国债利率

图 2-27　从 1962 年 1 月 4 日到 2009 年 4 月 10 日美国周利率的散点图

其中 $R^2=96.5\%$，两个系数的标准误差分别为 0.024 和 0.004．模型（式（2-46））证实了两种利率之间的高度相关性．然而，由图 2-28 所示的该模型残差时序图及残差 ACF 可见，该模型是严重不充分的．特别地，残差的样本 ACF 是高度显著的，并且缓慢衰减，显示出单位根非平稳时间序列的特点．残差的表现说明两种利率间存在明显的差异．利用现代计量经济学方法，如果我们假定两个利率序列都是单位根非平稳的，那么式（2-46）中残差的表现说明这两种利率序列不是**协整的**（cointegrated），见 Tsay，（2010，第 8 章）关于协整的讨论）．换言之，数据不支持关于两种利率间存在长期均衡关系的假设．从某种意义上讲，这不意外，因为“逆收益曲线”（inverted yield curve）在数据所在的时间段内出现了．“逆收益曲线”指的是利率与它们离到期日的时间长短逆相关．

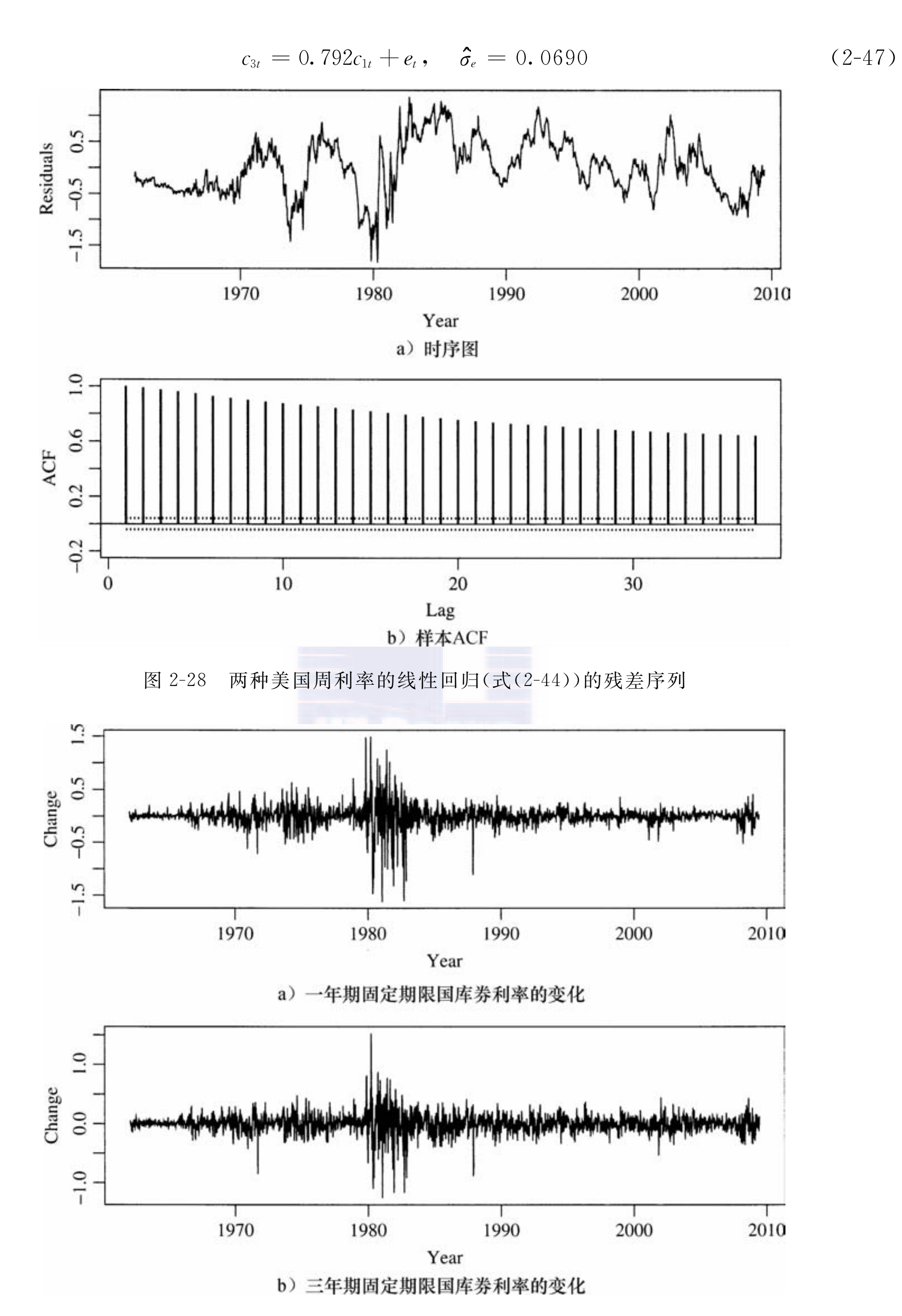
两个利率序列和式（2-46）中的残差的单位根表现使我们想到考虑利率变化量序列．令

1）$c_{1t}=x_{1t}-x_{1,t-1}=(1-B)x_{1t}$，$t\geqslant 2$：一年期利率的变化量．

2）$c_{3t}=x_{3t}-x_{3,t-1}=(1-B)x_{3t}$，$t\geqslant 2$：三年期利率的变化量．

并考虑线性回归 $c_{3t}=\beta c_{1t}+e_t$．图 2-29 所示的是两个变化量序列的时序图，而图 2-27b 所示的是它们之间的散点图．这两个变化量序列仍然是高度相关的，对它们拟合的线性回归模型为

$$c_{3t} = 0.792 c_{1t} + e_t, \quad \hat{\sigma}_e = 0.0690 \tag{2-47}$$



a）时序图

b）样本ACF

图 2-28　两种美国周利率的线性回归(式(2-44))的残差序列

a）一年期固定期限国库券利率的变化

b）三年期固定期限国库券利率的变化

图 2-29　从 1962 年 1 月 12 日到 1999 年 9 月 10 日美国周利率变化量序列的时序图

其中 $R^2=82.5\%$，系数的标准误差是 0.0073. 此模型进一步证实了两种利率间很强的线性相关性. 图 2-30 给出了式(2-47)残差的时序图和样本 ACF. 该样本 ACF 又表明残差中有一些显著的前后相关性，但相关系数的绝对值很小. 残差的这种弱前后相关性能用前面几节中讨论的简单时间序列模型来描述，从而得到一个带时间序列误差的线性回归.

这一节的主要任务是讨论用一个简单方法来建立带时间序列误差的线性回归模型. 这种方法是直接的. 我们对残差序列选用一个本章所讨论过的简单时间序列模型，然后联合估计整个模型. 为了说明，考虑式(2-47)中的简单线性回归. 因为模型的残差是前后相关的，所以我们给残差识别一个简单的 ARMA 模型. 从图 2-30 的样本 ACF 看出，MA(1) 模型对残差序列是适合的，把线性回归模型修改为

$$c_{3t}=\beta c_{1t}+e_t,\quad e_t=a_t-\theta_1 a_{t-1} \tag{2-48}$$

其中$\{a_t\}$是一个白噪声序列. 换句话说，我们简单地用不带常数项的 MA(1)模型来刻画式(2-47)中误差项的前后相关性. 结果得到的模型是带时间序列误差的线性回归模型的简单例子. 在实际应用中，可把更复杂的时间序列模型加到线性回归方程上去，形成通用的带时间序列误差的回归模型.

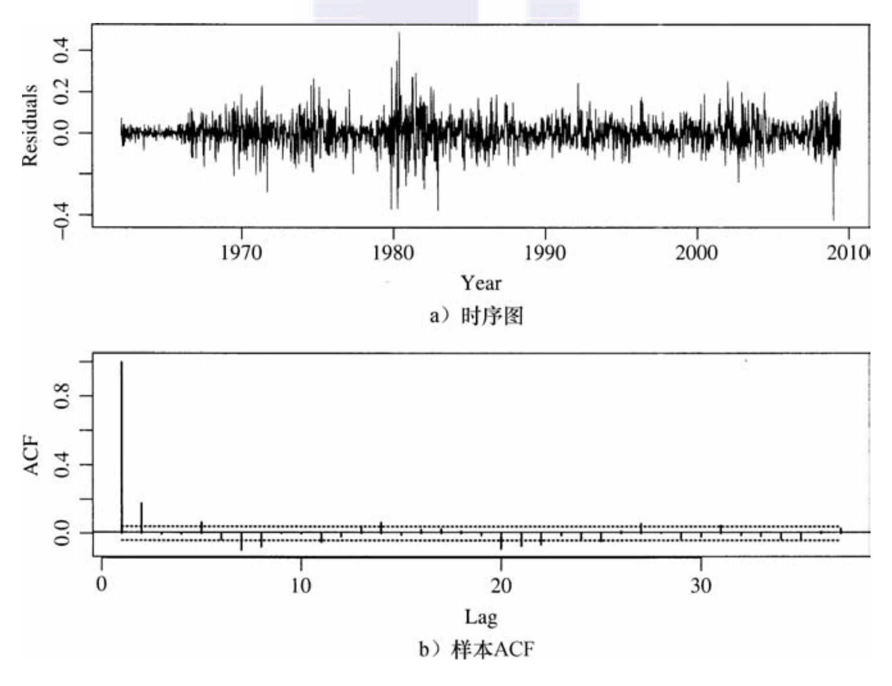


图 2-30　两种美国周利率变化的线性回归(式(2-45))的残差序列

在现代计算机出现之前，估计带时间序列误差的回归模型不是一件容易的事. 人们提出一些特殊的方法，如 Cochrane-Orcutt 估计量，来处理残差中的前后相关性(Greene，2003，第 273 页). 但现在这种估计与其他时间序列模型的估计一样容易. 若所用的时间序列模型是平稳的、可逆的，则可用最大似然法把模型一起估计出来. 这就是我们应用 R 的命令 `arima` 时所应用的方法. 对美国的周利率数据，所拟合的模型(式(2-48))为

$$c_{3t}=0.794c_{1t}+e_t,\quad e_t=a_t+0.1823a_{t-1},\quad \hat{\sigma}_a=0.0678 \tag{2-49}$$

111~114

其中 $R^2=83.1\%$，参数的标准误差分别为 0.0075 和 0.0196. 此模型不再有显著的滞后阶数为 1 的残差 ACF，虽然在滞后阶数为 4、6 和 7 时残差有较小的前后自相关性，但是，如果在残差方程中加上滞后阶数为 4、6 和 7 的 MA 系数，结果改进并不大，这里没有给出具体结果.

比较式(2-46)、式(2-47)和式(2-49)这三个模型，我们观察到如下几点：第一，模型(式(2-46))的高 R^2(96.5%)和系数(0.930)导致人们的错误认识，因为该模型的残差有强烈的前后相关性. 第二，对于利率变化量序列，式(2-47)和式(2-49)对应模型的 R^2 和 c_{1t} 的系数都很接近. 这说明对这个具体例子来说，给变化量序列加上一个 MA(1)模型所得到的改进不大. 这一点不出乎意料，因为估计出的 MA 系数不是较大的数值，尽管它是高度统计显著的. 第三，上述分析表明在线性回归分析中检验残差的前后相关性是很重要的.

从式(2-49)，该模型表明前述两个周利率序列有如下关系：

$$x_{3t}=x_{3,t-1}+0.794(x_{1t}-x_{1,t-1})+a_t+0.182a_{t-1}$$

这两种利率是相互影响并前后相关的.

R 代码演示(输出经过编辑)

```
> r1=read.table("w-gs1yr.txt",header=T)[,4]
> r3=read.table("w-gs3yr.txt",header=T)[,4]
> m1=lm(r3~r1)
> summary(m1)
Call:
lm(formula = r3 ~  r1)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.83214    0.02417   34.43   <2e-16 ***
r1           0.92955    0.00357  260.40   <2e-16 ***
---
Residual standard error: 0.5228 on 2465 degrees of freedom
Multiple R-squared: 0.9649,     Adjusted R-squared: 0.9649

> plot(m1$residuals,type='l')
> acf(m1$residuals,lag=36)
> c1=diff(r1)
> c3=diff(r3)
> m2=lm(c3~-1+c1)
> summary(m2)
Call:
lm(formula = c3 ~ -1 + c1)
Coefficients:
   Estimate Std. Error t value Pr(>|t|)
c1 0.791935   0.007337   107.9   <2e-16 ***
---
Residual standard error: 0.06896 on 2465 degrees of freedom
Multiple R-squared: 0.8253,     Adjusted R-squared: 0.8253

> acf(m2$residuals,lag=36)

> m3=arima(c3,order=c(0,0,1),xreg=c1,include.mean=F)
> m3
```

```
Call:
arima(x = c3, order = c(0, 0, 1), xreg = c1, include.mean = F)
Coefficients:
         ma1      c1
      0.1823  0.7936
s.e.  0.0196  0.0075

sigma^2 estimated as 0.0046: log likelihood=3136.62, aic=-6267.23
>
> rsq=(sum(c3^2)-sum(m3$residuals^2))/sum(c3^2)
> rsq
[1] 0.8310077
```

小结

我们给出分析带时间序列误差的线性回归模型的一般步骤：

1）拟合一个线性回归模型并检验其残差的前后相关性.

2）如果残差序列是单位根非平稳的，则对因变量和解释变量都做一阶差分. 然后对两个差分后的序列进行第 1 步. 若这时的残差序列是平稳的，则对残差识别一个 ARMA 模型并相应地修改线性回归模型.

3）用最大似然法进行联合估计，并对模型进行检验，查看是否需要进一步改进.

为检验残差的前后相关性，我们推荐使用 Ljung-Box 统计量，而不使用 Durbin-Watson(DW)统计量，因为后者只考虑滞后阶数为 1 的前后相关性. 有时残差的前后相关性表现在高阶间隔上，尤其是在所涉及的时间序列呈现某种季节性的时候.

注记：设有残差序列 e_t 的 T 个观察值，Durbin-Watson 统计量是

$$\mathrm{DW}=\frac{\sum_{t=2}^{T}(e_t-e_{t-1})^2}{\sum_{t=1}^{T}e_t^2}$$

直接计算表明 $\mathrm{DW}\approx 2(1-\hat{\rho}_1)$，其中 $\hat{\rho}_1$ 是 $\{e_t\}$ 的滞后阶数为 1 的 ACF. ■

注记：在一些应用中，人们可能着重于对线性回归模型做出推断，而不是直接对残差序列的相关性结构感兴趣. 这时候，得到回归系数的协方差矩阵的相合估计就足够了. 在出现相关性和条件异方差时，文献中有多个方法可以用来估计回归系数的协方差矩阵. 感兴趣的读者可以参阅 Tsay(2010，2.10 节). ■

115 ~ 116

2.11 长记忆模型

我们已经讨论过，平稳序列的 ACF 在滞后阶数增加时呈指数衰减. 但是，对单位根非平稳时间序列，可以证明对任意固定的滞后阶数，当样本容量增加时，样本 ACF 收敛于 1 (Chan 和 Wei(1988)；Tiao 和 Tsay(1983)). 另外也存在一些时间序列，随着滞后阶数的增加它们的 ACF 以多项式的速度缓慢衰减到 0. 这些时间序列称为**长记忆时间序列**(long-memory time series). 长记忆时间序列的一个例子就是如下定义的**分数差分序列**：

$$(1-B)^d x_t = a_t \quad -0.5<d<0.5 \tag{2-50}$$

其中$\{a_t\}$是一个白噪声序列. 模型(式(2-50))的性质已在文献中得到广泛的研究(Hosking, 1981). 我们把它的一些性质概括如下：

1) 若 $d<0.5$，则 x_t 是弱平稳过程并有无穷阶 MA 表示：

$$x_t = a_t + \sum_{i=1}^{\infty} \psi_i a_{t-i}$$

其中

$$\psi_k = \frac{d(1+d)\cdots(k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!}$$

2) 若 $d>-0.5$，则 x_t 是可逆的并有无穷阶 AR 表示：

$$x_t = \sum_{i=1}^{\infty} \pi_i x_{t-i} + a_t$$

其中

$$\pi_k = \frac{-d(1-d)\cdots(k-1-d)}{k!} = \frac{(k-d-1)!}{k!(-d-1)!}$$

3) 对$-0.5<d<0.5$，x_t 的 ACF 为

$$\rho_k = \frac{d(1+d)\cdots(k-1+d)}{(1-d)(2-d)\cdots(k-d)}, \quad k=1,2,\cdots$$

特别地，$\rho_1=d/(1-d)$，且当 $k\to\infty$时

117

$$\rho_k \approx \frac{(-d)!}{(d-1)!} k^{2d-1}$$

4) 对$-0.5<d<0.5$，x_t 的偏自相关函数(PACF)为 $\phi_{k,k}=d/(k-d)$，$k=1$，2，….

5) 对$-0.5<d<0.5$，作为 x_t 的 ACF 的傅里叶变换的 x_t 的谱密度函数 $f(\omega)$，它满足

$$f(\omega) \sim \omega^{-2d}, \quad \omega \to 0 \tag{2-51}$$

其中 $\omega\in[0, 2\pi]$表示频率.

当 $d<0.5$ 时，我们对 x_t 的 ACF 的性质特别感兴趣. 这个性质表明 $\rho_k\sim k^{2d-1}$是以多项式速度衰减，而不是以指数速度衰减. 正因为这个原因，这样的 x_t 过程称为长记忆时间序列. 式(2-51)中的谱密度的特殊特征是，当 $\omega\to 0$ 时谱发散到无穷. 然而，平稳 ARMA 过程的谱密度函数对所有 $\omega\in[0, 2\pi]$是有界的.

前面我们用到了非整数幂的二项式定理：

$$(1-B)^d = \sum_{k=0}^{\infty}(-1)^k \binom{d}{k} B^k, \quad \binom{d}{k} = \frac{d(d-1)\cdots(d-k+1)}{k!}$$

若分数差分序列$(1-B)^d x_t$ 服从一个 ARMA(p, q)模型，则称 x_t 为一个 ARFIMA(p, d, q)过程，它是 ARIMA 模型的推广，这里允许 d 为非整数.

在实际中，如果一个时间序列的样本 ACF 在数值上不大，但衰减得很慢，则该序列可能有长记忆. 作为说明，图 2-31 所示的是 1970 年 1 月 2 日至 2008 年 12 月 31 日的 CRSP 价值加权指数和等权重指数的日简单收益率的绝对值序列的样本 ACF 值. 可见 ACF 的数值相对较小，但衰减得很慢，甚至在滞后阶数为 300 之后还是在 5%的水平下是显著的. 对于绝对值收益率序列的样本 ACF 的表现，更多的讨论参见 Ding 等(1993). 对式(2-50)中的纯分数差分模型，我们可以用最大似然法或带低频对数周期图的回归方法来估计 d.

可以参考，例如 Geweke 和 Porter-Hudak(1982)的工作．最后，长记忆模型在金融文献中受到关注，部分原因是由于在连续时间模型中关于分形布朗运动的工作．

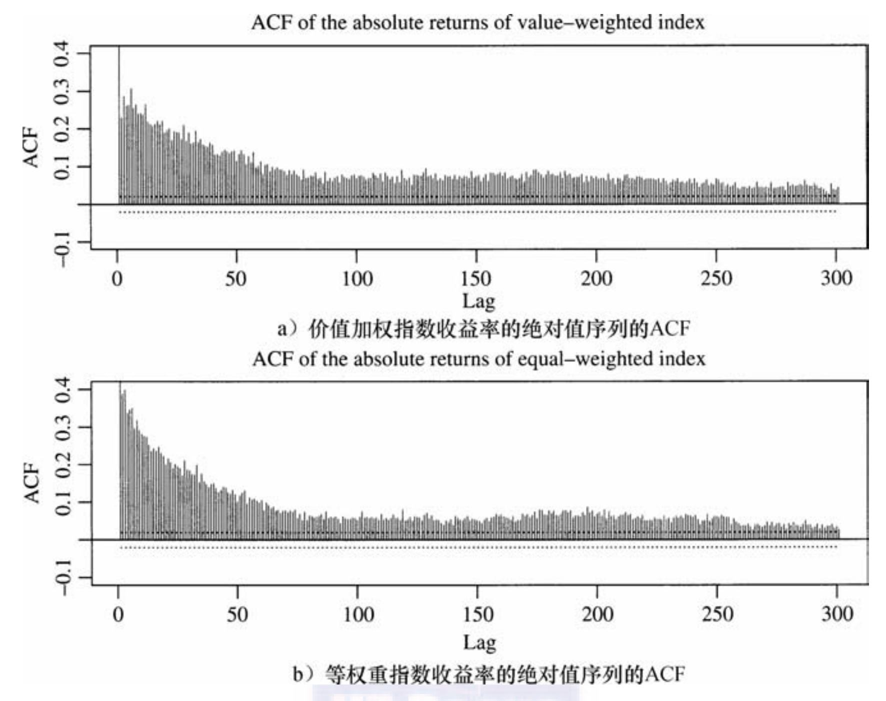


图 2-31　CRSP 价值加权指数和等权重指数的日简单收益率的绝对值序列的样本 ACF．样本周期为 1970 年 1 月 2 日到 2008 年 12 月 31 日

例 2.8　为了说明长记忆时间序列的建模，我们考虑 CRSP 价值加权指数从 1970 年 1 月 2 日到 2008 年 12 月 31 日的绝对日收益率．图 2-31a 给出了序列的样本 ACF 图．把该序列作为纯分数差分过程并应用 Geweke-Porter-Hudak 方法，我们得到 $\hat{d}=0.372$，标准误差为 0.070．该估计值的确在平稳和可逆的区间(0，0.5)内．如果我们应用 AFRIMA(1，d，1)模型，并应用最大似然估计法，我们得到拟合的模型为

$$(1-0.113B)(1-B)^{0.491}y_t=(1-0.576B)a_t$$

其中 y_t 是绝对收益，估计的标准误差按照出现顺序分别为 0.006、0.008 和 0.006．这些估计是高度显著的．特别地，分数参数的估计接近非平稳的边界 0.5．最后，可以应用 R 的 `fracdiff` 添加包来估计 AFRIMA(p，d，q)模型．■

R 代码演示

```
> library(fracdiff)
> da=read.table("d-ibm3dx7008.txt",header=T)
> head(da)
      Date       rtn     vwretd     ewretd     sprtrn
1 19700102  0.000686  0.012137  0.033450  0.010211
 ....
6 19700109 -0.001353 -0.002797 -0.002923 -0.003021
> ew=abs(da$vwretd)
% obtain Geweke-Port-Hudak estimate using command fdGPH
```

```
> m3=fdGPH(ew)
> m3
$d
[1] 0.3722260
$sd.as
[1] 0.0698385
$sd.reg
[1] 0.06868857
% Maximum likelihood estimation of an AFRIMA(1,d,1) model.
> m2=fracdiff(ew,nar=1,nma=1)
> summary(m2)
Call:
  fracdiff(x = ew, nar = 1, nma = 1)

Coefficients:
   Estimate Std. Error z value Pr(>|z|)
d  0.490938   0.007997   61.39   <2e-16 ***
ar 0.113389   0.005988   18.94   <2e-16 ***
ma 0.575895   0.005946   96.85   <2e-16 ***
---
[d.tol = 0.0001221, M = 100, h = 0.0003742]
Log likelihood: 3.551e+04 ==> AIC = -71027.02 [1 deg.freedom]
```

2.12 模型比较和平均

在实际应用中，对于一个给定时间序列来说完全正确的模型是不存在的．所有的统计模型都只能估计数据的动态相关性，而且适合给定数据的拟合模型不止一个．这样，就会产生多个模型如何比较的问题．本节中，我们将介绍比较时间序列模型的两种方法．这两种方法都属于统计准则，它们提供了模型选择的一些规则．然而，它们也只是一些参考规则，事实上，在选择模型时数据分析的目的和所考虑问题的实际信息都是要考虑的重要因素．

2.12.1 样本内比较

如果数据分析的目的是为了研究一个时间序列的动态结构，那么我们可以用样本内方法来比较不同的模型．样本内法就是利用所有数据来进行模型估计和比较．这样，信息准则(如 AIC 和 BIC)和残差方差的估计都可用来进行模型比较．如果选定其中一个准则，那么它的值越小，模型就越好．例如，考虑 CRSP Decile 1 指数从 1970 年 1 月到 2008 年 12 的数据，我们利用例 2.7 中的季节 ARMA 模型和 1 月为哑变量的回归模型．对于该回归模型，残差的标准误差是 0.069，而在季节 ARMA 模型中该值为 $\sqrt{0.00472}=0.0687$．如果样本内比较法把残差的标准误差作为模型选择标准，那么就会选择季节 ARMA 模型．对于这个例子，这两个模型间的差别很小．

118 ~ 120

2.12.2 样本外比较

若建立时间序列模型是为了预测，那么进行模型比较就要考虑模型的预测能力．在进行样本外比较时，一般用预测误差的均方(Mean Square of Forecast Error，MSFE)来量化模型的预测能力．这种模型比较的方法在金融文献中叫做**回测检验**(backtesting)．我们利

用超前 1 步预测来介绍这种方法．不过，该方法对超前多步预测也适用．

回测检验

统计模型回测检验的步骤如下：

1）将数据集分割为估计子样本和预测子样本．这里没有具体的指导准则，但是每一个子样本要包含足够多的数据点，确保估计和 MSFE 的计算尽可能准确．

2）用估计子样本中的数据进行估计，利用拟合模型得到超前 1 步预测和它的预测误差．具体地，我们假设估计子样本是$\{x_t|t=1,\cdots,h\}$．用前 h 个数据点来计算超前 1 步预测$\hat{x}_h(1)$和它的预测误差 $e_h(1)=x_{h+1}-\hat{x}_h(1)$．模型估计过程中没有用到数据点 x_{h+1}．

3）对估计子样本添加一个新的数据点，变为$\{x_t|t=1,\cdots,h+1\}$．用前 $h+1$ 个数据点重新估计模型，计算超前 1 步预测和它的预测误差．也就是，计算 $e_{h+1}(1)=x_{h+2}-\hat{x}_{h+1}(1)$，其中$\hat{x}_{h+1}(1)$是以 $h+1$ 为预测原点的新拟合模型的超前 1 步预测．

4）重复步骤 3），直到超前 1 步预测误差为 $e_{T-1}(1)=x_T-\hat{x}_{T-1}(1)$，$T$ 为样本量．

那么，模型的 MSFE 是

$$\mathrm{MSFE}(m)=\frac{\sum_{j=h}^{T-1}[e_j(1)]^2}{T-h}$$

其中 m 表示所用的模型．我们把最小 MSFE 对应的模型作为这组数据的最好模型．在实际中，我们用 MSFE 的平方根，而不是 MSFE 本身．度量模型预测能力的其他度量有平均绝对预测误差(Mean absolute Forecast Error，MAFE)和偏差(Bias)，它们分别定义为：

$$\mathrm{MAFE}(m)=\frac{\sum_{j=h}^{T-1}|e_j(1)|}{T-h},\quad \mathrm{Bias}(m)=\frac{\sum_{j=h}^{T-1}e_j(1)}{T-h}$$

121

例 2.9　考虑从 1947 年第 1 季度到 2010 年第 2 季度的美国季度实际 GDP．GDP 数据来源于圣路易斯(St Loius)联邦储备银行(Federal Reserve Bank)，已做了季度调整，以 2005 年 GDP 为基础进行通胀调整，以 10 亿美元为单位(billions of chained 2005 dollars)．图 2-32 表明

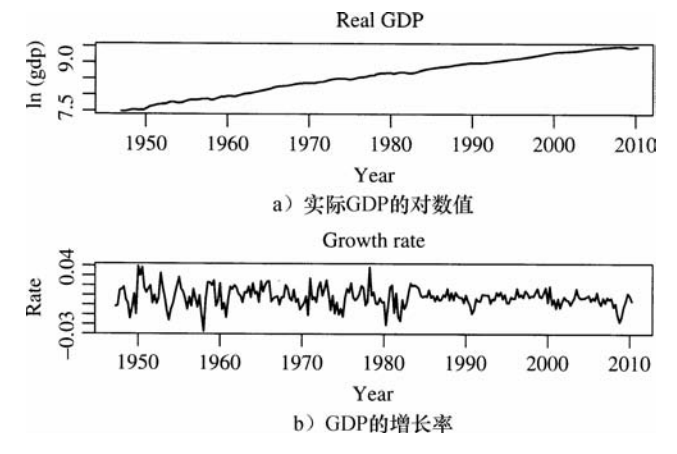


图 2-32　从 1947 年第 1 季度到 2010 年第 2 季度美国实际季度 GDP 的时序图

实际季度 GDP 的对数值和增长率序列．我们关注的是增长率序列，记为 x_t．图 2-33 显示了 x_t 的样本 ACF 和 PACF．从这个图可以看出，它们衰减的速度很快．PACF 说明该数据可以用 AR(3)模型．利用 R 语言中的 ar 命令获取的 AIC 值也表明，AR(3)模型较好．拟合的模型为：

$$(1-0.346B-0.130B^2+0.123B^3)(x_t-0.0079)=a_t,\quad \hat{\sigma}_a^2=8.32\times 10^{-5} \quad (2\text{-}52)$$

其中估计的标准误差依次分别是 0.062、0.066、0.062 和 0.0009．拟合的 AR(3)模型的 AIC 是 −1648.45．图 2-34 表明 AR(3)模型的模型检验统计量．这些图可以说明 AR(3)模型是充分的．

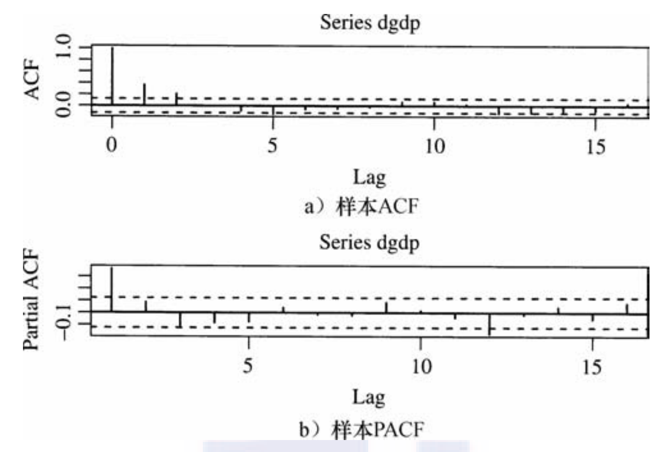


图 2-33 美国从 1947 年第 2 季度到 2010 年第 2 季度的季度实际 GDP 增长率的样本 ACF 和 PACF．虚线表示两个标准误差的上下限

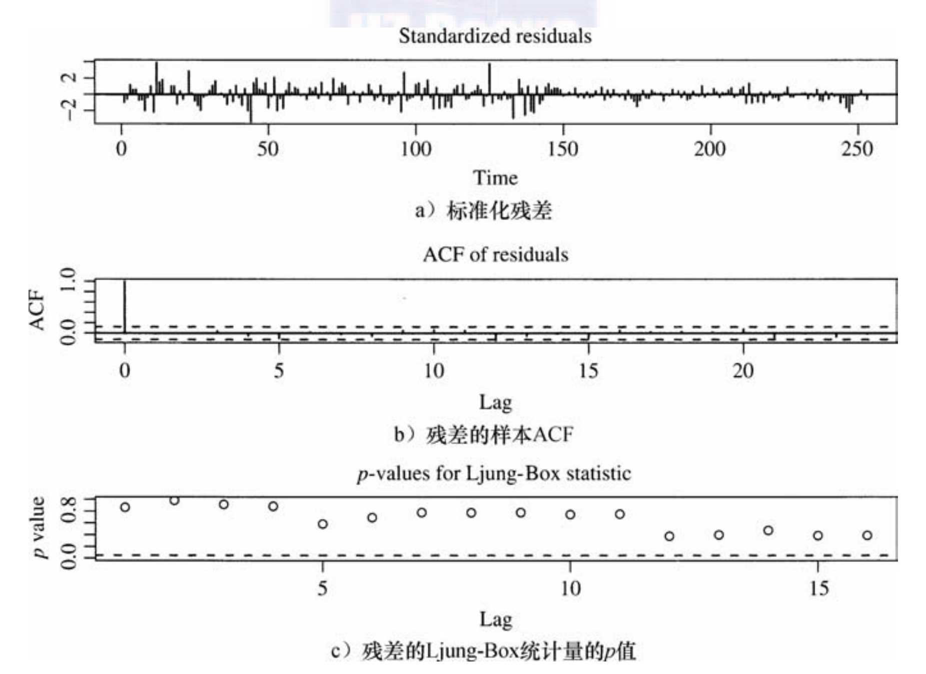


图 2-34 从 1947 年第 2 季度到 2010 年第 2 季度美国的实际季度 GDP 增长率的拟合 AR(3)模型的检验

由于数据已做了季节调整，所以为了进行模型比较，我们也使用季节模型

$$(1-0.331B-0.152B^2+0.110B^3)(1-0.497B^4)(x_t-0.0079)=(1-0.587B^4)a_t \quad (2\text{-}53)$$

其中估计的标准误差分别是 0.063、0.067、0.064、0.258、0.0008 和 0.236，残差方差是 $\tilde{\sigma}_a^2=8.24\times10^{-5}$. 季节 AR 参数仅仅略微显著. 模型检验不能说明拟合的季节模型不充分性. 模型(式(2-53))的 AIC 是－1646.93. ■

现在，我们对式(2-52)和式(2-53)的模型进行比较. 用样本内比较法，AIC 选择 AR(3)模型. 用样本外法进行比较，以 2000 年第四季度为预测原点，按照回测检验的步骤，在预测时间段内有 38 个季度. 对于超前 1 步预测来说，AR(3)模型和季节模型的预测误差的平方根分别为 0.006 15 和 0.006 32，这说明 AR(3)模型更好. 平均绝对预测误差分别是 0.004 43 和 0.004 55，这也说明 AR(3)模型比季节模型好.

122 ～ 123

最后，上例中，对于实际 GDP 增长率序列来说，样本内和样本外法都选择了同一个 AR(3)模型. 在实际应用中，样本内和样本外法可能会选择不同的模型，初始预测原点的选取也可能影响模型的选择结果. 因此，我们建议选取不同的初始预测原点进行比较，以便更好地理解模型比较的结果.

注记： 例 2.9 的回测检验是用 R 脚本 backtest 来实现的，可以在本书的网站下载. ■

R 代码演示

```
> da=read.table("q-gdpc96.txt",header=T)
> head(da)
  Year Mon Day      gdp
1 1947   1   1 1772.204
 ....
6 1948   4   1 1856.930
> gdp=log(da$gdp)
> dgdp=diff(gdp)

> m1=ar(dgdp,method='mle')
> m1$order
[1] 3
> m2=arima(dgdp,order=c(3,0,0))
> m2
Coefficients:
         ar1     ar2      ar3  intercept
      0.3462  0.1299  -0.1225     0.0079
s.e.  0.0623  0.0655   0.0624     0.0009

sigma^2 estimated as 8.323e-05: log likelihood=829.23, aic=-1648.45

> m3=arima(dgdp,order=c(3,0,0),season=list(order=c(1,0,1),period=4))
> m3
Coefficients:
         ar1     ar2      ar3    sar1     sma1  intercept
      0.3305  0.1521  -0.1103  0.4966  -0.5865     0.0079
s.e.  0.0633  0.0668   0.0635  0.2578   0.2357     0.0008
```

```
sigma^2 estimated as 8.24e-05: log likelihood=830.47, aic=-1646.93
>
> source("backtest.R")    % Perform backtest
> mm2=backtest(m2,dgdp,215,1)
[1] "RMSE of out-of-sample forecasts"
[1] 0.006153102
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.004430387
> mm3=backtest(m3,dgdp,215,1)
[1] "RMSE of out-of-sample forecasts"
[1] 0.006322009
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.004553896
```

2.12.3 模型平均

当有多个模型可以很好地拟合给定的时间序列数据时，我们除了从中选择某个模型外，也可以利用所有这些模型产生一个组合预测．这种技术在统计文献中叫做模型平均(model averaging)．假设有 m 个可用模型，它们都能给出一个时间序列的无偏预测．这里无偏预测是指相关联的预测误差的期望为 0．令 $\hat{x}_{i,h+1}$ 是模型 i 在预测原点 h 的超前 1 步预测，那么组合预测可定义为

$$\hat{x}_{h+1} = \sum_{i=1}^{m} w_i \hat{x}_{i,h+1}$$

其中 w_i 是一个非负实数，表示模型 i 的权重，满足 $\sum_{i=1}^{m} w_i = 1$．权重 w_i 可以通过多种方式来确定．比如，在贝叶斯推断中，w_i 是模型 i 的后验概率．这里，我们采用简单平均，即 $w_i = \frac{1}{m}$．有限的经验表明，这种简单平均在实际中很有效．

习题

除非特别声明，在以下习题中都用 5%的显著性水平来得出结论．

1. 考虑从 1948 年 1 月到 2011 年 11 月美国失业率的月数据(见文件 m-unrate-4811.txt)，数据来自美国圣路易斯的联邦储备银行．
   (a) 该失业率的月数据是否存在单位根？为什么？
   (b) 根据该数据建立一个时间序列模型并检验模型是否已充分拟合数据．然后，根据所建立的模型对美国 2011 年 12 月和 2012 年前 3 个月的失业率进行预测．(注意：适合该数据的模型不止一个，只要模型充分即可．)
   (c) 拟合的模型是否存在商业周期？为什么？

124 ~ 125

2. 以 NYSE/AMEX/NASDAQ 的市场资本为基础考虑 CRSP Decile 1、2、5、9、10 投资组合的月简单收益率．该数据的时间区间是从 1961 年 1 月到 2011 年 9 月．
   (a) 对于 Decile 2 和 Decile 10 的收益序列，在 5%的显著性水平下检验：原假设是滞后阶数为 1～12 的自相关系数均为 0．给出你的结论．
   (b) 对于 Decile 2 的收益率序列建立一个 ARMA 模型，对模型进行检验并写出拟合的模型．
   (c) 利用拟合的 ARMA 模型对序列进行超前 1～12 步预测，并给出预测的相关标准误差．

3. 考虑在 2007 年 1 月 2 日到 2011 年 12 月 23 日期间，苹果公司股票每天的股价波动幅度(即当天的最高价减去当天的最低价). 这个数据可以用 R 添加包 quantmod 从雅虎财经获得. 计算这个时间序列开始 100 滞后期数的 ACF 值，存在长范围相依的证据吗？为什么？如果这个数据存在长记忆，请建立该数据的 ARIMA 模型.
4. 考虑 Moody 公司 Aaa 和 Baa 级季度债券从 1919 年 1 月到 2011 年 11 月的月收益率. 该数据源于美国圣路易斯联邦储备银行的 FRED. 考虑 Aaa 债券的月对数收益率序列，建立时间序列模型并对模型进行检验.
5. 继续考虑 Aaa 债券的月对数收益序列，利用指数平滑法求超前 1～12 步的样本外预测，预测原点是 2010 年 11 月.
6. 考虑前面所讨论的两个债券的收益率序列，它们之间存在什么关系？为了回答这个问题，首先对数据取对数，然后利用 Baa 级债券的收益率作为解释变量，建立 Aaa 级债券收益率数据的时间序列模型. 写出拟合的模型并对模型进行检验.
7. 考虑强生(Johnson&Johnson)公司股票从 1992 年第 1 季度到 2011 年第 2 季度每股季度收益数据. 文本文件 q-jnj-earns-9211.txt 中的数据来源于 Thomson Reuters 的 Fisrt Call Historical Database. 如果需要可以对数据取对数. 建立时间序列模型并进行模型检验，写出所拟合的模型. 利用 1992 年到 2008 年的数据重新进行拟合，并写出季度收益的超前 1～10 步预测，画出预测值的时序图.
8. 考虑美国从 1947 年第 1 季度到 2011 年第 3 季度的季度实际 GNP，该数据存放于文件 q-GNPC96.txt 中，数据已做了季节调整，以 2005 年 GNP 为基础进行了通胀调整，以 10 亿美元为单位(billions of chained 2005 dollars). 假设 x_t 代表 GNP 增长率的时间序列数据.
   (a) 通过 ar 命令，应用 AIC 准则，可以为 x_t 识别一个 AR(4)模型. 拟合这个模型，拟合的模型充分吗？为什么？
   (b) 数据 x_t 的样本 PACF 识别的是 AR(3)时间序列模型. 拟合这个模型，拟合的模型充分吗？为什么？
   (c) 如果用样本内法进行比较，哪个模型好？为什么？
   (d) 以 2004 年第 4 季度为预测原点，把数据分为估计样本和预测样本，应用回测检验，并以 MSFE 为标准，为 x_t 选择一个模型，并给出选择该模型的理由.

126

参考文献

Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. 2nd International Symposium on Information Theory. Budapest: Akademia Kiado; 1973. p 267–281.

Box GEP, Jenkins GM, Reinsel GC. Time Series Analysis: Forecasting and Control. 3rd ed. Englewood Cliffs (NJ): Prentice Hall; 1994.

Box GEP, Pierce D. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J Am Stat Assoc 1970; 65: 1509–1526.

Brockwell PJ, Davis RA. Introduction to Time Series and Forecasting. 2nd ed. New York: Springer; 2002.

Brockwell PJ, Davis RA. Time Series: Theory and Methods. 2nd ed. New York: Springer; 2009.

Chan NH, Wei CZ. Limiting distributions of least squares estimates of unstable autoregressive processes. Ann Stat 1988; 16: 367–401.

Cryer JD, Chan KS. Time Series Analysis: With Applications in R. New York: Springer-Verlag; 2010.

Dickey DA, Fuller WA. Distribution of the estimates for autoregressive time series with a unit root. J Am Stat Assoc 1979; 74: 427–431.

Ding Z, Granger CWJ, Engle RF. A long memory property of stock returns and a new model. J Empir Finance 1993; 1: 83–106.

Fuller WA. Introduction to Statistical Time Series. 2nd ed. Hoboken (NJ): John Wiley & Sons; 1995.

Geweke J, Porter-Hudak S. The estimation and application of long memory time series models. J Time Anal 1982; 4: 221–238.

Greene WH. Econometric Analysis. 5th ed. Upper Saddle River (NJ): Prentice-Hall; 2003.

Hosking JRM. Fractional differencing. Biometrika 1981; 68: 165–176.

Ljung G, Box GEP. On a measure of lack of fit in time series models. Biometrika 1978; 66: 67–72.

Peña D, Tiao GC, Tsay RS. A Course in Time Series Analysis. Hoboken (NJ): Wiley & Sons; 2001.

Phillips PCB. Time series regression with a unit root. Econometrica 1987; 55: 277–301.

Shumway RH, Stoffer DS. Time Series Analysis and its Applications. New York: Springer; 2000.

Tiao GC, Tsay RS. Consistency properties of least squares estimates of autoregressive parameters in ARMA models. Ann Stat 1983; 11: 856–871.

Tsay RS. Analysis of Financial Time Series. 3rd ed. Hoboken (NJ): Wiley & Sons; 2010.

Tsay RS, Tiao GC. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. J Am Stat Assoc 1984; 79: 84–96.

Woodward WA, Gray HL, Elliott AC. Applied Time Series Analysis. Baca Raton (FL): CRC Press, Taylor & Francis Group; 2012.

127